Can Icons Outperform Text? Understanding the Role of Pictograms in OHMD Notifications

Nuwan Janaka
nuwanj@comp.nus.edu.sg
National University of Singapore
NUS-HCI Lab, School of Computing
Singapore

Shengdong Zhao
zhaosd@comp.nus.edu.sg
National University of Singapore
NUS-HCI Lab, School of Computing
Singapore

Shardul Sapkota
shardul@u.yale-nus.edu.sg
National University of Singapore
NUS-HCI Lab, School of Computing
Singapore

ABSTRACT

Optical see-through head-mounted displays (OHMDs) can provide just-in-time digital assistance to users while they are engaged in ongoing tasks. However, given users' limited attentional resources when multitasking, there is a need to concisely and accurately present information in OHMDs. Existing approaches for digital information presentation involve using either text or pictograms. While pictograms have enabled rapid recognition and easier use in warning messages and traffic signs, most studies using pictograms for digital notifications have exhibited unfavorable results. We thus conducted a series of four iterative studies to understand how we can support effective notification presentation on OHMDs during multitasking scenarios. We find that while *icon-augmented notifications* can outperform text-only notifications, their effectiveness depends on icon familiarity, encoding density, and environmental brightness. We reveal design implications when using *icon-augmented notifications* in OHMDs and present plausible reasons for the observed disparity in literature.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Mobile devices**; *Information visualization.*

KEYWORDS

notification, OHMD, OST HMD, smart glasses, pictogram, icon, interruption, distraction

ACM Reference Format:
Nuwan Janaka, Shengdong Zhao, and Shardul Sapkota. 2023. Can Icons Outperform Text? Understanding the Role of Pictograms in OHMD Notifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3544548.3580891

1 INTRODUCTION

Optical see-through head-mounted displays (OST HMD, augmented reality smart glasses, OHMD) [48] allow users to access information anytime, anywhere. However, in mobile multitasking scenarios—such as when attending to notifications while walking— attention resources allotted to process the digital information presented on OHMDs are rather limited [71]; thus, it is essential for the information presented to be concise and intuitive so that users can quickly and accurately understand their meaning at a glance. These requirements motivate an important research question: How can digital information on OHMDs be presented in a way to support effective information acquisition under multitasking scenarios?

Current approaches to digital information presentation involve using either text (verbal) or pictograms (graphical). While text is typically the default choice, pictograms have been used in a number of non-digital scenarios to allow rapid recognition, provide concise information, and support easier usage [27, 93, 100, Ch 6]. For example, by converting part of the warning messages into pictograms, traffic road signs became universally understandable at a glance [27, 52]. Such benefits have been observed in several other application domains: such as tracking pilot vehicles [13] and adhering to health warnings [41, 58].

These successful examples have encouraged researchers to investigate further the potential benefits of incorporating pictograms, either by themselves or mixed with text, for digital information display on computing devices (e.g., notifications [88, 98]). Nevertheless, till today, most studies have not revealed favorable results, as incorporating pictograms in digital notifications has either exhibited no additional benefits or performed worse than using text only. For example, Warnock et al. [98, 99] have found that there is no difference between texts and pictograms on users' primary task performance and distraction levels. Subsequent studies conducted by Tanveer et al. [90] and Somervell et al. [88] have shown that, in fact, users preferred text over pictorial feedback.

This disparity in digital notifications has piqued our interest and motivated us to investigate this issue further. To ground our investigations in a practical application domain, this study investigated whether it is possible to enhance the users' digital notification reading performance on OHMDs, while they are engaged in a primary task by replacing part of the text notification with commonly used pictograms (icons) in the form of *icon-augmented notifications*. Specifically, we scoped our exploration on calendar notifications in stationary and mobile situations due to their high frequency of occurrence (e.g., more than 8% of total daily mobile notifications [82]) and high perceived value such as for reminding events or to-do actions and acting as memory-aids [51, 82, 95]. We conducted a series of four iterative studies (three laboratory studies and one in-the-wild study), and our results showed that *icon-augmented notifications* could outperform text-only notifications, yet their effectiveness depends on several factors, including users' familiarity

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580891

with the icon (i.e., frequency of use and intuitiveness of the depicted object [46]), encoding density (i.e., the number of words represented by the icon), and environmental contexts (e.g., external lighting). The findings of this study help to shed light on the reasons behind the observed inconsistent results in the literature and provide insights on when and how *icon-augmented notifications* can be used on OHMDs. Finally, we discuss how these results can be generalized to other types of notifications, i.e., *icon-augmented notifications* are suitable when limited secondary information has to be conveyed, and well-designed icons can be found to represent the primary information.

The contributions of this paper are threefold: 1) identifying factors affecting the effectiveness of *icon-augmented notifications* in near-eye displays when users are engaged in a primary task; 2) enhancing the understanding of how text and pictogram format used for OHMD notifications differ in terms of interruption, reaction, and comprehension (IRC framework); and 3) suggesting design implications when applying *icon-augmented notifications* in specific domains (calendar events) and realistic settings (mobile scenarios).

2 RELATED WORK

This research focuses on two goals, 1) enhancing our understanding of the potential role of *icon-augmented notifications* in multitasking OHMD usage, and 2) identifying the plausible reasons for the observed disparity in the previous studies. The related work is organized into three sub-categories related to these two goals: first, an overview of notification management research, is provided, then previous investigations related to notification design on OHMDs are discussed, and finally, the disparity observed in the literature related to the effectiveness of using pictograms vs. text in various contexts is examined.

2.1 Notification management

Although notifications increase the intake of digital information, enable proactive communication, reinforce a sense of connectedness, and help with task management [39, 45, 73, 75], they often reduce task performance, increase mental load, generate negative emotions, and disrupt social interactions [1, 7, 23, 40, 54, 57, 62, 89].

To balance access to notification information while minimizing attention costs [64], researchers proposed several strategies, which can be broadly classified into three categories: *mediating* strategies, which defer notifications until the user is more receptive to them; *indicating* strategies, which indicate the availability of the receiving party to the sending party; and, *mitigating* strategies, which change the device or presentation (modality) of the notifications to make them less distracting [2].

Of all the above strategies, *mitigating* strategies are the only type that provides notification information in a timely manner which can be particularly useful if immediate attention is required. Instead of delaying message appearance, *mitigating* strategies reduce distraction by re-encoding messages into an easier-to-understand representation to reduce the cognitive load. This paper investigates a particular mitigating strategy which is re-encoding partial textual information into icons.

2.2 Notifications on OHMDs

Similar to other device notifications, OHMD notifications tend to interrupt the users' primary tasks and reduce their work performance [59, 62]. Hence, to minimize interference with primary tasks when presenting information on OHMD, previous research has focused on different mitigating strategies, including the use of peripheral vision [16, 22, 47, 49, 60], adjusting presentation timing [70, 90], adjusting presentation modality [20, 24, 31, 59, 70, 90], and adjusting content placement [19, 53, 80]; with the latter including: content stabilization [33, 53, 56], alignment [80], and position [19, 53, 80]. In that regard, Klose et al. [53] and Fukushima et al. [33] explored content stabilization mechanisms by locking content through world, body, and head anchoring, and found that text readability improves with world and body anchoring; however, head anchoring is preferred for urgent texts such as notifications.

With regard to presentation modality, most of the works have focused on using multiple sensory modalities (e.g., visual vs. audio) [20, 24, 31, 70], while only a few have focused on using a single sensory modality [59, 90]. Consequently, this paper focuses on a single sensory modality, particularly the visual presentation modality, as it can be combined with other sensory modalities in a complementary manner [77] and is the predominant output mode in OHMDs [48]. Concerning a single sensory modality, previous work includes Lucero and Vetek's 'NotifEye' system [59], where they used animated butterflies as a cue for incoming notifications and allowed users to attend OHMD notifications with subtle interactions; however, their focus was on the pre-presentation stage of the notifications. In comparison, this study focuses on the presentation of the notification itself and seeks to reduce distraction and preserve communicative effectiveness by modifying the presentation modality using pictograms.

2.3 Pictograms in notifications

Pictograms, such as graphical illustrations, icons, and signs, are commonly used in daily navigation and digital applications in computing devices. They are also widely used in general communication, such as emojis/emoticons in messages [91, 105], communication in emergency scenarios [32, 103], and for people with intellectual disabilities [85].

Despite the prevalent usage of pictograms, there is no consensus on their effectiveness compared to the use of text. A large body of work, especially in the domain of traffic sign research (for reviews, see [8, 15]), favors the use of pictograms as they are not dependent on language [93], allow rapid recognition [27, 52], present information concisely [13, 66]. These advantages were demonstrated in a number of studies that found that pictograms were easier to use than verbal/text equivalents [14, 93].

On the contrary, another set of previous studies argues against the above notion [44, 87, 92, 101]. For instance, Warnock et al. [98, 99] investigated different notification modalities, including text and pictograms, in a desktop computing context for delivering home care reminders for older people who were engaged in a card-matching game; and the authors found that disruption during the game did not vary with the modality of the notifications. Moreover, Roca et al. [79] compared traffic signs in single words with pictograms and found that single-word messages were associated with better performance (farther reading distances) and required less visual demands (fewer glances and less glancing times) than pictograms. In addition, in the case of OHMDs, Tanveer et al. [90] delivered feedback on OHMDs during public speaking and found that the feedback in the form of text was more effective and easier to learn than pictorial feedback. As a result, using pictograms alone is considered less effective in communicating information as compared with the use of text [101].

Given this disparity in the literature about the effectiveness of using pictograms, we are interested in further investigating this problem to uncover the plausible reasons for such results.

3 *ICON-AUGMENTED NOTIFICATIONS* ON OHMDS

In exploring the reasons for differences between pictograms vs. text, we ground our investigation in a practical HCI problem: how to design effective notifications on OHMDs for multitasking usage. In particular, we are interested in whether the incorporation of *icons* in OHMD notifications can minimize attention costs (e.g., distraction, task interference). Although current mobile notifications use icons to display the notification source (e.g., app, sender) in a supplementary manner, the content of the notification is still entirely displayed using the text format (e.g., Figure 1a) [3, 5]. In contrast, our work explores how we can represent the content of the notifications itself partially via icons (e.g., Figure 1b), which is different from existing approaches.

Icons are symbolic signs of objects and concepts and are widely used visual representations in user interfaces and traffic signs [14, 93]. Icons are graphical, thus easier to recognize and remember [93, 100, Ch 6], but unlike pictures— which can lead to a wide range of interpretations [92]— each icon is typically designed to be associated with one meaning [14]. In this sense, it is functionally similar to logographical words (e.g., Chinese characters). A recent study conducted by Huang et al. [42] using brain imaging has shown that icons are not cognitively processed as logographical words (although they both stimulate the semantic system in the brain that is needed for language processing) but are more similar to images and pictures. Thus, icons can take advantage of our brain's unique capabilities to process images. Icons can be used either alone or mixed with text. From previous research, we learned that icons alone could be difficult to interpret [101]; thus, we chose the alternative approach to mix icons with text to create *icon-augmented notifications* and study whether this particular form of notifications can have any advantages over text alone.

3.1 *Icon-augmented notification* design

While there are many kinds of notifications, in this investigation, we focused on calendar notifications as their high frequency in daily life ($M \approx 4$, $SD > 16$ notifications per day according to Table 2 in [82]), uniform structure (sec 3.1.1), and high perceived value [51, 82, 95].

3.1.1 Notification structure. Calendar notifications could be thought of as being composed of two parts: 1) primary information (e.g., event/action) and 2) secondary information (e.g., time, person, or location). The notification "Meeting in 30 minutes" can thus be decomposed into "<*primary info*: meeting, *secondary info*: 30 minutes>". To convert the text notification into *icon-augmented notification*, *primary info* was represented using icons, while *secondary info* was represented using numbers or text. An example is shown in Figure 1c. Note that for conciseness, prepositions, such as the word 'in' from the notification "Meeting in 30 minutes" were removed, and abbreviations were used in the *icon-augmented notification*, as our pilot studies have shown that they do not affect comprehension.

3.1.2 Icon selection and design calibration. To reduce the gap of familiarity, we choose commonly used icons from Google Material Icons[1] and Flaticon website (premium license)[2]. We use outline style as it has been shown that on OHMDs, outline styles are preferred as it permits more viewing of the environment, which enhances multitasking (i.e., navigation) performance [76].

To ensure a fair comparison, three human raters, a designer, a researcher, and a proofreader, independently evaluated the *text notifications* and their corresponding *icon-augmented notifications* for similarity in information difference and intuitiveness. Twenty-four calendar notifications, adapted from real notifications, were designed in *text notification* and corresponding *icon-augmented notification* sets (e.g., Figure 1c) after two iterations until raters reached full consensus. See Appendix A.1 for details.

3.1.3 Notification layout. As recommended by Debernardis et al. [26], all texts on OHMD were displayed in green color with a sans serif font (Roboto[3]). Our pilot study showed that text with a font of 50 sp[4] under equal heights, an icon size of 50sp x 50sp, presented the best combination for space utilization and clarity on our OHMD device (Figure 2, sec 5.2). Notifications were displayed on the top-center position, as recommended by Chua et al. [19] for multitasking situations where primary tasks require central attention.

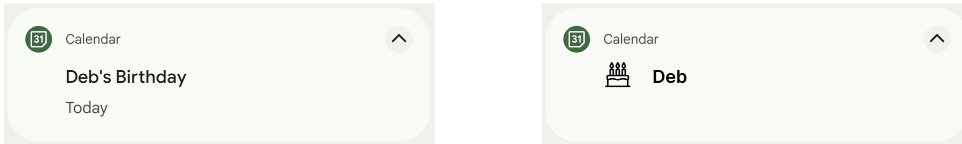
4 RESEARCH APPROACH AND OVERVIEW

This research aims to achieve two goals: one practical and one theoretical. The former goal concerns whether the new design of *icon-augmented notifications* can outperform the text-based notifications in OHMD multitasking scenarios, while the latter aims to explain the disparity observed in the notification literature about the effectiveness of pictograms in notifications. Hence, this research starts with a study to address the first goal and iteratively probes the problem space based on the empirical findings.

4.1 *Study 1*: Compare *text notifications* vs *icon-augmented notifications* for *researcher-selected* icons

In the first study (sec 5), the design of *icon-augmented notifications* was compared to **regular** *text notification* in a controlled multitasking scenario. The results showed that *icon-augmented notifications* outperform the regular *text notifications* in terms of task performance and reduced distraction. This finding provides

[1] Material design icons - https://material.io/resources/icons/?style=outline
[2] Flaticon website - https://www.flaticon.com/
[3] https://fonts.google.com/specimen/Roboto
[4] *sp* represents the *scalable pixels* which are the same as *dp (density-independent pixels)* for default text size, https://developer.android.com/training/multiscreen/screendensities



(a) (b)

regular *text notification*	*<primary info, secondary info>*	*icon-augmented notification*
Meeting at 4 pm	<meeting, 4 pm>	 4 pm
Doctor's appointment in 2 hours	<doctor appointment, 2 hrs>	 2 hrs

(c)

Figure 1: A comparison between *text notifications* and *icon-augmented notifications*. (a) A typical calendar notification [3, 5] where the content is fully represented by text. (b) The proposed *icon-augmented notification* where the content is partially represented via icons. (c) Examples of *text notification* to *icon-augmented notification* mapping. Icon sources: Flaticon website (premium license)

(a) (b)

Figure 2: The notification layout on OHMD. The notifications are displayed at the top-center position. (a) A *text notification* on OHMD. (b) An *icon-augmented notification* on OHMD. Note: Black color in OHMD represents the transparent background. Icon source: Flaticon website (premium license).

evidence that the proposed *icon-augmented notifications* can be advantageous over *text notifications*; however, this doesn't address the second goal, as a careful examination of the study design revealed a potential confounding variable, filler words (e.g., linking words between *primary info* and *secondary info*, see Figure 1c). Due to the different affordances between icon-augmented and text-based notifications, filler words were retained in *text notifications* but not in *icon-augmented notifications*, which resulted in participants spending additional time reading regular *text notifications*. Thus, this confounding variable makes it difficult to draw any conclusion regarding the effectiveness of pictogram vs text.

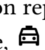
4.2 *Study 2*: Compare *icon-augmented notifications* and transformed *text notifications* for *researcher-selected* icons

To eliminate the potential bias, the filler words were removed from the *text notifications*, and a second study was conducted (sec 6, see Figure 5). Before conducting this study and to ensure that the resulting *text notifications* are comprehensible, a pilot study with 4 participants was conducted, and the results showed that the removal of filler words in the *text notifications* does not significantly affect their comprehension.

Examining the results of the *study 2*, it was observed that the previously observed statistical advantages in *icon-augmented notifications* from *study 1* have disappeared, although the majority of participants still subjectively preferred *icon-augmented notifications*. The results of this study indicated that other than subjective feelings, replacing text with icons does not bring any statistically significant advantage.

While this finding is aligned with past studies that showed no advantages of pictograms over text, it can't explain why other previous studies have found advantages in pictograms. Hence, by carefully examining the design of the first two studies, two additional potential influencing factors were identified: **content familiarity** and **encoding density**.

Regarding the former factor, due to the difference in training and practice, participants were more familiar with text stimuli as they have been using text daily for many years; on the other hand, participants had less exposure to the icons used in the experiment and were not familiar with the meanings of all icons. Although we tried to narrow the gap with training and practice before the experiment, our post-experimental feedback revealed that participants were still not familiar with all icons used in the experiment, and such differences could potentially bias the results.

Regarding the latter, encoding density, which refers to the number of words an icon represents, is another potential influencing factor. For example,  represents 'Car', which is one word, but  represents 'No Left Turn', which is three words. Arguably, the higher the density, the more efficient the encoding process in the brain, which can affect the effectiveness of recognizing the notifications.

4.3 *Study 3*: Compare *icon-augmented notifications* with transformed *text notifications* for *user-selected* icons

To understand how these two factors may affect the results obtained from *study 2*, a third study (sec 7, see Figure 7) was conducted by improving the participants' familiarity with the selected icons and adding encoding density as another independent variable. In particular, instead of providing a predefined set of icons as stimuli, participants were allowed to choose their own icons, which is believed to help in increasing the familiarity with the stimuli.

The results of the *study 3* showed that: 1) with increased familiarity, *icon-augmented notifications* have again demonstrated statistical advantage over *text notifications*, and 2) an interaction effect between presentation *format* × *encoding density* was found. A detailed analysis showed that the observed advantages in *icon-augmented notifications* are mostly attributed to the higher *density* conditions but not the lower ones.

Combining the findings of the three studies, a plausible explanation behind the reasons for the inconsistent results observed in the literature can be deduced; this explanation can be related to the fact that when comparing the performance of pictogram and text, many factors come into play. Pictograms can outperform text in multitasking performance, but with several conditions, which are: 1) the design of the pictogram needs to be intuitive and unambiguous, 2) users need to be highly familiar with the pictogram, and 3) the *density* of the pictogram needs to be high. If any of the conditions are not met, the advantages of pictograms may not be observed.

4.4 *Study 4*: Compare *icon-augmented notifications* and transformed *text notifications* in realistic settings for *user-selected* icons

Examining the performance of *icon-augmented notifications* on OHMDs in real life can achieve a better understanding of how pictograms compare with text. Thus, a fourth study (sec 8) was conducted in realistic stationary and mobile settings to verify the generalizability of laboratory findings.The results largely confirmed the findings from the laboratory studies, where *icon-augmented notifications*, when carefully designed, can reduce distraction as compared to text. However, it also revealed additional factors that can influence users' performance, such as the amount of external lighting.

The above is an overview of the current research, and the next few sections will present each of the above studies in detail.

4.5 Common setting

All three controlled studies (*Study 1, 2, and 3*) were based on a dual-task scenario [72] to validate whether adapting pictorial representation reduces attention costs while retaining the content delivery. A *vigilance task* (i.e., an attention-demanding task, a perceptual monitoring task) was chosen as the primary task for which attentional control was measured [78, 84], and users were instructed to attend to the calendar notifications as the secondary task. Similar approaches have been used to measure the attentional cost of mobile phone notifications [89] and multitasking on OHMDs [53, 68].

5 STUDY 1: COMPARE *TEXT NOTIFICATIONS* VS *ICON-AUGMENTED NOTIFICATIONS* FOR *RESEARCHER-SELECTED* ICONS

In this study, *text notifications* and *icon-augmented notifications* were compared with the *no-notification* condition, and the task performance as well as the user's preference, were evaluated.

5.1 Participants

Sixteen volunteers (8 females, 8 males, mean age = 22.7 years, SD = 2.5) participated in this study. All participants had normal or corrected to normal visual acuity with no reported color or visual deficiencies/impairments. All participants were from the university community who had self-reported a minimum level of professional working fluency in English and were avid smartphone users who received on average 51 (min = 30, max = 100) notifications per day; yet, none had prior experiences with OHMDs.

For **all** studies, the participants consented to the experiments and were compensated ≈ USD 7.25/h for their time. **None** of the participants in any study were involved in the subsequent studies.

5.2 Apparatus

The study was conducted in a quiet room under indoor lighting conditions to provide a consistent user experience and avoid confounding factors due to environmental interference [26, 34].

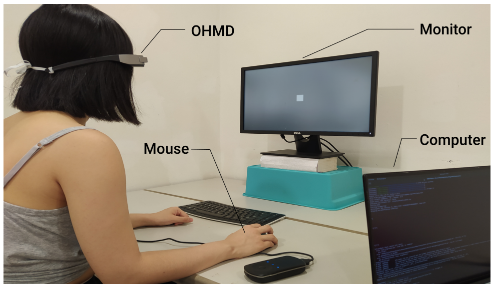
The primary task was displayed on a light grey background on a 23" LCD monitor (refresh rate = 60 Hz, resolution = 1920 x 1080 px) at eye level (see Figure 3a) and was designed using PsychoPy[5] [74]. The stimuli of the primary task and notifications were presented at different depths to simulate attention switching between physical and virtual backgrounds. Hence, the participants' eyes were set to be 70 cm away from the computer monitor, which differs from the focal length of around 1m [55] used by OHMDs. Moreover, the notifications were presented in an Epson Moverio BT-300 smart glasses [28] (Figure 3b), a binocular OHMD with 1280x720 px (30Hz) resolution display, 23° FoV, and a projected distance of 80 inches at 5m running on Android 5.1 OS (headset weight = 69 g). A BT-300 was selected as our OHMD since its functionality/features are a subset of the more advanced OHMDs (such as HoloLens2, Nreal Light, etc.); thus, its results could be better generalized to a wide range of OHMDs (details at sec 9.2.5).

In addition, a custom-developed Android application was installed on the OHMD to display custom notifications (Figure 2). A Python program handled the monitor's displayed stimuli, pushing notifications to the OHMD, logging user inputs, and synchronizing timings (see Appendix E for implementation details).

5.3 Tasks

For the *primary (vigilance) task*, we adopted the shape detection task developed by Santangelo et al. [83] and reused by Mustonen et al. [68] to evaluate the visual task performance of an OHMD. This vigilance task was chosen for two reasons. First, the vigilance task is an established representative task that can simulate real-world

[5] A Python library used for experimental psychology research



(a) Apparatus setup

(b) Epson BT-300 smart glasses

Figure 3: Study apparatus used in the controlled experiments

usage of OHMDs in dynamic and unpredictable environments [68]. This is particularly relevant in situations where attention needs to be divided between the display (virtual content: notifications) and the environment (physical content: shape detection), such as walking in crowded areas. Moreover, this type of task can measure the degradation of sustained visual attention and perceptual monitoring capabilities when receiving virtual content, which affects gaze, attention, and situational awareness [68, 69]. Second, a controlled experimental task is needed for a fair comparison. Although real-world tasks have higher external validity, they often have many confounding variables that are hard to control.

During the shape detection task, the visual stimuli continuously morph between small (15 x 15 mm^2) and large (30 x 30 mm^2) white squares for 625 ms in segments that last 3750 ms. While the aforementioned small and large white squares are non-targets (Figure 4), a target shape that is either a vertical (15 x 30 mm^2) or horizontal (30 x 30 mm^2) rectangle randomly appears in 88.9% of segments, and no two rectangles appeared within 1875 ms of each other. Participants were instructed to press the left mouse button upon detecting a target rectangle shape at a time limit of 1875 ms, and the total duration of the task was four minutes to ensure sufficient time for assessing participants' attention control [68].

The *secondary (notification) task* was to attend to the calendar notifications on the OHMD. Six notifications were randomly displayed, during the primary task, with a minimum interval of 20 seconds between each notification and a display duration of 10 seconds, similar to the study conducted by Rzayev et al. [80]. Calendar notifications designed in sec 3.1 were used in this task.

5.4 Study Design and Procedure

A repeated-measures within-subject design was used to investigate the participants' performance on primary and secondary tasks for three notification presentation *formats*: *text notification*, *icon-augmented notification*, and *no-notification*, with the latter being used as the comparative baseline. The experiment consisted of three testing blocks, each with a duration of four minutes, which were counterbalanced using a Latin square.

5.4.1 Procedure. Participants were first instructed to familiarize themselves with the *researcher-selected* icon-to-text mapping (sec 3.1.1).

Next, two verbal recognition tests were administered on OHMD to verify the participants' icon recognition accuracy for all icons. The notifications related to the icons that participants recognized wrongly in the second test were removed before the training and testing blocks to minimize the effects of unfamiliar/unintuitive icons. The resulting notifications (22-24 per participant) were used by the apparatus (sec 5.2) to show notifications randomly without repetition.

Afterward, a training session for at least two minutes each in all conditions was conducted until participants felt comfortable with the apparatus, tasks (primary and secondary), and questionnaires. Then, the participants underwent three testing blocks knowing the condition, and were instructed to attend to the primary task as quickly and accurately as possible. At the end of each block, participants filled out a questionnaire that recorded their perceived behaviors and recalled notifications. A minimum break of two minutes with eye exercises was given between blocks to minimize fatigue.

Upon completion of all blocks, participants filled out a questionnaire with their overall rankings for each format and attended an 8-12 minutes semi-structured post-interview in which they were asked about the reasons for each ranking, the process, and the multitasking experience with different formats. Each experiment comprised one session lasting 40-55 minutes.

5.5 Measures

5.5.1 Primary (vigilance) task. Both accuracy and speed are measured for the primary task. To measure the accuracy, *hit rate* ($H = \frac{\#hit}{\#hit + \#miss} \in [0, 1]$, hit = correct identification of target shape) and *false alarm rate* (F[6] $= \frac{\#false\ alarm}{\#false\ alarm + \#correct\ rejection} \in [0, 1]$, $false\ alarm$ = misidentifying a noise signal as a target signal) were used; while to measure speed, *reaction time* (RT= response time - target stimuli start time, in *seconds*) was used. A failure to respond within the time limit was considered a miss, and reaction times were calculated only for the hits.

5.5.2 Secondary (notification) task. Given the attention-utility trade-off of notifications, McCrickard et al. developed the *IRC framework* focusing on user goals to evaluate notification systems [63–65]. Their model used three critical parameters: *Interruption*: an event prompting transition and reallocation of attention focus from a task to the notification, *Reaction*: rapid and accurate response to the notification stimuli, and *Comprehension*: remembering and making sense of the information that the notifications conveyed at a later time, which captures the multifaceted nature of notifications. These parameters were used to evaluate the proposed *icon-augmented*

[6]Since $target$ % = 88.9% > $noise$ % = 10.1%, the precision of F is lower than that of H

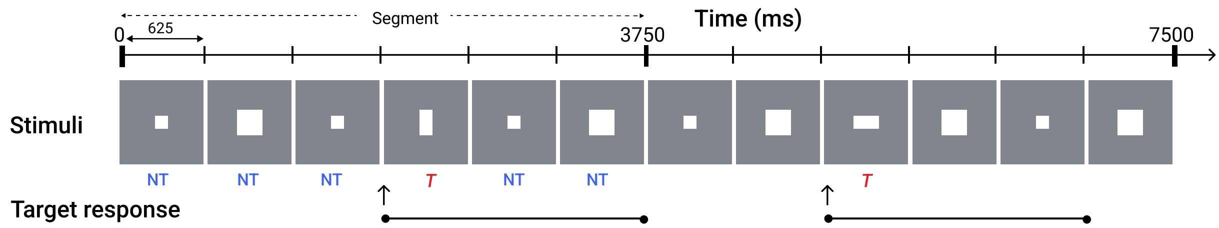


Figure 4: The stimulus is a square (non-targets = NT) that morphs between small and large sizes in cycles of 625 ms. The stimulus morphed into the target shape (T = vertical or horizontal rectangle) at random, and participants were instructed to respond by pressing their mouse button within 1875 ms. This figure depicts the square shapes morphing to target shapes in 4^{th} morph in 1^{st} segment (0-3750 ms) and 3^{rd} morph in 2^{nd} segment (3750-7500 ms). Note: stimuli are not drawn to scale.

notifications by using operationalized measures [17]; i.e., perceived cost of interruption for *Interruption*, noticeability for *Reaction*, understandability and recall accuracy for *Comprehension* (see Table 1). Additionally, we considered *Satisfaction*: overall approval of notification, which was operationalized using preference [64]. As McCrickard et al. [63–65] described, the importance of individual parameters depends on the situational factors (e.g., context, user characteristics, information characteristics) as well as the user's goals. Therefore, we considered all parameters when analyzing our experiments to understand the multi-faceted nature of the proposed *icon-augmented notifications*. For instance, if the notification is about a reminder for an event that is happening in a few days, *Comprehension* would be the most important parameter to evaluate its effectiveness; if it is about an important meeting occurring in a few minutes that the user does not the recall at the moment, *Interruption* would be the most important criteria; lastly, if it is a notification that requires the user to take immediate action, then *Reaction* would be the most important parameter.

Table 1: Measures based on IRC framework [17, 64, 65]. Here [O] represents objective measures while [S] represents subjective measures.

Parameter	Measures
Interruption	[S] Perceived cost of interruption (task load)
Reaction	[S] Immediate response (noticeability)
Comprehension	[O] Base comprehension (recall accuracy and understandability)
Satisfaction	[S] Preference

5.5.3 Interruption. Perceived task load using raw NASA-TLX (*RTLX*, 0-100 scale) [38], *Perceived Interruption* ('How much interruption did the notification cause to the task when you attempt to carry out both simultaneously?') using a 0-100 visual analogue scale, and distraction ranking were used to measure the perceived cost of interruption.

5.5.4 Comprehension. Immediate recall accuracy (*Recall Accuracy*) and understandability ranking were used to measure the comprehension of each *format*. *Recall Accuracy* was calculated for the notifications displayed while participants were engaged with the primary task, using a questionnaire after each block. *Recall Accuracy* was used in this study in order to simulate the scenario where certain calendar notifications need to be remembered to take due action in the future [51, 95]. For each correct *primary info*, 0.5 points were assigned, and another 0.5 points were assigned to correct *secondary info* when *primary info* was correct (see Appendix A.2 for examples).

5.5.5 Reaction. Noticeability ranking was measured to understand how *format* affects the rapid detection of notifications.

5.5.6 Satisfaction. Overall preference ranking was used to measure the desirability of each *format*.

5.6 Study 1: Results

Each participant completed three blocks and received 12 notifications and 192 targets. They scored a minimum of 3 (out of 6) for recall accuracy and more than 72% hit rate at the end of each notification block. One participant's data was removed as an outlier since the hit rate deviated more than three times the standard deviation away from the mean. Table 2 indicates participants' mean performance of measures.

5.6.1 Analysis. A one-way repeated measures ANOVA was used to analyze the quantitative data when the baseline data were present; in the cases where ANOVAs assumptions were violated, the Friedman test was used. When there was no baseline data, a paired-sample t-test or Wilcoxon signed-rank test was used. The normality of the data was tested using the Shapiro-Wilk test, and the sphericity was tested using the Mauchly test. Finally, as suggested by Huberty and Morris [43], statistical tests were conducted on each dependent variable separately as they are conceptually distinct aspects, and parametric tests were used when non-parametric distributions could take a large range of values and follow parametric assumptions.

As for the interview recordings, they were transcribed and analyzed thematically following the process outlined by Braun and Clarke [11], and the qualitative findings were grouped into themes to support/oppose quantitative findings.

5.6.2 Primary (vigilance) task performance. As expected, both notification *formats* significantly reduced the primary task accuracy

Table 2: Performance in dual-task scenario (N = 15). Colored bars show the relative value of each measure for different notification formats. ⋆ † ‡ represent significant post-hoc tests ($p < 0.05$). Here, *Text = text notification, Icon = icon-augmented notification, No = no-notification, H = hit rate, F = false alarm rate, RT = reaction time,* and *RTLX = Raw NASA-TLX score.*

Measure	*H*		*F*		*RT*		*RTLX*		*Perceived Interruption*		*Recall Accuracy*	
Format	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
No	0.960†‡	0.042	0.052	0.108	0.464⋆	0.044	29.17⋆†	17.35	-	-	-	-
Icon	0.938⋆‡	0.055	0.073	0.147	0.479	0.039	48.33†	16.75	52.00⋆	21.36	4.83	0.88
Text	0.923⋆†	0.057	0.086	0.134	0.486⋆	0.050	55.17⋆	18.37	63.67⋆	20.91	4.80	0.75

(i.e., low hit rate, high false alarm rate) and speed (i.e., low reaction time), indicating that notifications degrade the primary task's performance.

However, *icon-augmented notification* had a higher hit rate, lower false alarm rate, and shorter reaction time than *text notification*, indicating that *pictogram format* enables it to sustain the primary task performance better than *text format*.

- *Hit rate*: Friedman test revealed a significant effect ($\chi^2(2) = 6.778$, p=0.035, W = 0.709, strong agreement[7]) of notification format. Pair-wise Wilcoxon signed-rank tests with Bonferroni correction indicated that hit rate for *no-notification* ($M = 0.960$, $SD = 0.042$) was significantly higher ($p_{bonf} < 0.05$) than *icon-augmented notification* ($M = 0.938$, $SD = 0.055$) and *text notification* ($M = 0.923$, $SD = 0.057$), and that *icon-augmented notification* was significantly higher ($Z = 69.5$, $p = 0.050$, $r = 0.527$, large effect[8]) than *text notification*.
- *False alarm rate*: There was no significant effect between the different notification formats.
- *Reaction time*: There was a significant effect of the notification's format ($F_{2,28} = 4.401$, $p = 0.022$, $\eta_p^2 = 0.239$, large effect[9]). A post-hoc analysis revealed that *text notification* ($M = 0.486$, $SD = 0.050$) was significantly different ($p_{bonf} < 0.05$) from *no-notification* ($M = 0.464$, $SD = 0.044$), but not significantly different from *icon-augmented notification* ($M = 0.479$, $SD = 0.039$).

5.6.3 Interruption. Although both *formats* significantly increased the cognitive load compared to *no-notification*, *icon-augmented notification* led to a lower cognitive load than *text notification*.

- *Unweighted NASA-TLX*: A repeated-measures ANOVA showed a significant effect ($F_{2,28} = 45.076$, $p < 0.001$, $\eta_p^2 = 0.763$, large effect) and a post-hoc analysis with Bonferroni corrections revealed that *no-notification* ($M = 29.17$, $SD = 17.34$) was significantly different ($p_{bonf} < 0.001$) from *text notification* ($M = 55.17$, $SD = 18.37$) and *icon-augmented notification* ($M = 48.33$, $SD = 16.75$), but *text notification* was not significantly different ($p_{bonf} = 0.069$) from *icon-augmented notification*. Results for individual indices are given in Appendix A.3.
- *Perceived interruption*: A paired-sample t-test showed that *text notification* ($M = 63.7$, $SD = 20.9$) was significantly more interruptive ($t(14) = 2.93$, $p = 0.006$, $d = 0.756$, medium effect[10]) than *icon-augmented notification* ($M = 52.0$, $SD = 21.4$).
- *Distraction ranking*: Eleven out of fifteen participants felt that *text notification* distracted their primary task the most as they often had to read the text over multiple glances, and it took them longer to absorb information as compared to the *icon-augmented notification*. On the other hand, three participants felt that the *icon-augmented notifications* were more distracting due to the lack of familiarity, while the remaining participant felt that the two formats had equal distraction levels.

5.6.4 Comprehension.

- *Immediate recall accuracy*: A Wilcoxon signed-rank test indicated that there was no significant difference between *icon-augmented notification* ($M = 4.83$, $SD = 0.88$) and *text notification* ($M = 4.80$, $SD = 0.75$) in terms of *Recall Accuracy*. This also indicates that the secondary (notification) task performance during multitasking was not affected by the *format*.
- *Understandability ranking*: Six (40%) participants felt that *text notifications* were more understandable than *icon-augmented notifications* as they did not need to interpret any icons, while six participants felt that *icon-augmented notifications* were easier to understand as they were shorter and easier to process mentally. The remaining three participants felt that both *formats* were equally understandable.

5.6.5 Reaction. Overall, there was no consensus on the noticeability of the two formats.

- *Noticeability ranking*: Seven (47%) participants felt that both *formats* were equally noticeable, while five participants felt that *icon-augmented notification* was more prominent since it is pictorially distinct. The remaining three participants felt that *text notification* was more noticeable due to their longer length.

5.6.6 Satisfaction. All participants, except for one (93%), preferred to receive *icon-augmented notifications* since icons were intuitive, shorter, caused less interruption, and were easier to recognize/interpret over long-term use.

[7] $0.3 \leq W < 0.6$ indicates moderate agreement and $0.6 \leq W$ indicates strong agreement [35, 94]
[8] $0.3 \leq r < 0.5$ indicates moderate effect and $0.5 \leq r$ indicates large effect [37, 94]
[9] $0.01 \leq \eta_p^2 < 0.06$ indicates small effect, $0.06 \leq \eta_p^2 < 0.14$ indicates medium effect, and $0.14 \leq \eta_p^2$ indicates large effect [37, 94]
[10] $0.2 \leq d < 0.5$ indicates small effect, $0.5 \leq d < 0.8$ indicates medium effect, and $0.8 \leq d$ indicates large effect [37, 94]

5.7 Discussion

The results show that *icon-augmented notifications* had a significantly higher primary task performance than *text notifications*. Hence, it can be deduced that the transformation facilitates multitasking since there was an increase in the primary task's performance while maintaining the secondary task performance (i.e., recall accuracy). Similarly, *icon-augmented notifications* significantly reduced *Interruption* while maintaining *Reaction* and *Comprehension*; additionally, they were preferred over *text notifications*. These results support the practical goal of this study (sec 4), the proposed *icon-augmented notifications* have advantages over *text notifications* during multitasking.

Nevertheless, it was observed that additional factors might affect the effectiveness of *icon-augmented notifications* based on the feedback from the participants. The presence of filler words in *text notifications* may have resulted in participants taking additional time to read the regular *text format*s as *icon-augmented notifications* used abbreviations while the regular-text version in *text notifications* did not. This feedback raised the question of whether the presence of filler words in *text notifications* affects its effectiveness during multitasking; thus, additional studies were carried out to examine the effect of filler words.

6 STUDY 2: COMPARE *ICON-AUGMENTED NOTIFICATIONS* AND TRANSFORMED *TEXT NOTIFICATIONS* FOR *RESEARCHER-SELECTED* ICONS

To control the effect of filler words and abbreviations, a second study was conducted to compare *icon-augmented notifications* with **transformed**, filler words removed (e.g., *meeting at 4 pm -> meeting 4 pm*), *text notifications* for *researcher-selected* icons. The same apparatus (sec 5.2) and task (sec 5.3) used in the previous study were used in this study.

6.1 Participants

Twelve volunteers (7 females, 5 males, mean age = 23.4 years, SD = 2.7) participated in this study. They had similar backgrounds to participants from *study 1* (sec 5.1), except for one participant who had previously used an OHMD for one hour.

6.2 Revised notification design

Following the same *notification design* procedure in *study 1* (sec 3.1.1), researchers selected icons to represent *primary info* and text/number for *secondary info*. To unify information content for the *icon-augmented notification* and *text notification*, the *icon-augmented notification* was transformed back to its text format without adding filler words (Figure 5). The three raters (who were in *study 1*) evaluated the *icon-augmented notification* and their **transformed** *text notification* to control for information difference and intuitiveness. If their content differed, the raters changed the *text notification* until full consensus was reached amongst the raters. Similarly, if any *icon-augmented notification* was not intuitive, the raters changed the icon.

6.3 Procedure

All aspects were similar to *study 1* (sec 5.4), except that the *no-notification* condition was removed during testing conditions. The *format* was fully counterbalanced using a Latin square.

6.4 Results

Participants scored a minimum of 2 (out of 6) for recall accuracy and more than 68% hit rate at the end of each notification block. Appendix B.1 indicates participants' mean performance of measures.

Surprisingly, there were no significant main effects or interaction effects for any quantitative measures ($p > 0.05$).

6.4.1 Primary (vigilance) task performance. In contrast to *study 1*, *pictogram format* and *text format* exhibit no significant difference (e.g., *hit rate* (pictogram: $M = 0.938$, $SD = 0.080$; text: $M = 0.949$, $SD = 0.040$, $p = 0.50$), *reaction time* (pictogram: $M = 0.481$, $SD = 0.049$; text: $M = 0.481$, $SD = 0.043$, $p = 0.62$)). These results indicated that when filler words were removed, the advantage of *pictogram format* disappeared as there was no significant difference in primary task measures.

6.4.2 Interruption. There was no significant difference in the perceived cost of interruption (e.g., *unweighted NASA-TLX* (pictogram: $M = 44.10$, $SD = 14.72$; text: $M = 45.66$, $SD = 13.55$, $p = 0.55$), *perceived interruption* (pictogram: $M = 47.92$, $SD = 20.22$; text: $M = 50.54$, $SD = 21.00$, $p = 0.50$)).

- *Distraction ranking*: In contrary to the quantitative measures, the qualitative feedback still showed that more participants (7/12) found *text format* more distracting than the *pictogram format*, quoting reasons similar to the ones that collected in *study 1* (sec 5.6.3), while the other participants thought the opposite; however, the difference was less than that observed in *study 1*.

6.4.3 Comprehension. Similar to *study 1*, there was no significant difference between *icon-augmented notification* ($M = 4.17$, $SD = 1.04$) and *text notification* ($M = 4.13$, $SD = 1.31$) in terms of *Recall Accuracy*.

- *Understandability ranking*: Slightly more participants (7/12) felt that the *text format* was easier to understand since their meanings were unambiguous, while three participants felt that the *pictogram format* was easier to understand since it was shorter and easier to process mentally, and the remaining two participants felt that both formats were similarly understandable.

6.4.4 Reaction. Similar to *study 1*, there was no consensus on the noticeability levels of the two formats. Six participants felt that noticeability levels were the same for both formats, while three participants felt that the *text format* was more noticeable, and the remaining three participants felt that the *pictogram format* was easier to notice.

6.4.5 Satisfaction. Interestingly, despite the lack of quantitative evidence, a majority of the participants (10/12) still preferred the *pictogram format*. Participants quoted similar reasons for their preferences as in *study 1* (sec 5.6.6).

regular *text notification*	*icon-augmented notification*		transformed *text notification*
Meeting at 4 pm		4 pm	Meeting 4 pm
Doctor's appointment in 2 hours		2 hrs	Doctor's appointment 2 hrs

Figure 5: Mapping from *icon-augmented notifications* to transformed *text notifications*. Icon source: Flaticon website (premium license).

6.5 Discussion

The results of this study revealed that the exhibit of the filler word is indeed a confounding variable. When filler words were removed, the performance between *pictogram format* and *text format* became overall comparable, indicating that text reduction influenced the increased multitasking performance and reduced distraction of *icon-augmented notification* in *study 1*. Yet, three caveats deserve additional attention. The first is that despite the lack of quantitative evidence support, the majority of the participants still preferred *pictogram format* over *text format*, indicating it is likely to still have some advantage (perhaps psychological). The second caveat is related to identifying a potential confounding variable, content familiarity, through analyzing the participants' feedback. While all participants were familiar with the text content, not all participants were familiar with the (researcher-selected) icons. First, participants mentioned that less familiar icons were confusing and harder to recall, especially if they were not relevant to their personal lives. Second, some icons contained similar symbols, which confused the participants. Thus, the different familiarity among the participants towards the *pictogram format* and *text format* stimuli could affect their performance. The third caveat is related to the encoding capabilities of the icons, which affect the distraction from notifications, *"when icons represent long sentences such as 'valentine day', it is way easier [to focus on the primary task] than text"*. In other words, it was easier to recognize *icon-augmented notifications* when icons captured more words than the *text notifications*.

7 STUDY 3: COMPARE *ICON-AUGMENTED NOTIFICATIONS* WITH TRANSFORMED *TEXT NOTIFICATIONS* FOR *USER-SELECTED* ICONS

In this study, the effects of content familiarity were further controlled by allowing participants to select their own icons (sec 7.2) and then compared *icon-augmented notifications* with transformed *text notifications*. Additionally, the number of words that each icon can represent was taken into account in this study by adding encoding density as an additional independent variable.

7.1 Participants

Twenty-four volunteers (11 females, 13 males, mean age = 23.7 years, SD = 3.8) participated in this study. They had similar backgrounds to the participants of *study 1* (sec 5.1), except for three participants who had in the past used an OHMD for approximately three hours.

7.2 Revised notification design

Since an icon can represent/encode either a single word (e.g., <birthday icon> represents 'birthday') or multiple words (e.g., <medical icon> to represent 'doctor's appointment'); hence, in the context of this paper, encoding density (*density*) is defined as the number of words an icon represents. Accordingly, an icon representing one word ("birthday") has an encoding density of 1, while an icon representing two words ("doctor's appointment") has an encoding density of 2, etc. In this study, *density* has two levels: *one* and *two*.

Three human raters, two designers and one co-author, chose four representative icons with outline style (see Figure 6) for *primary info* from Google Material Icons, Flaticon website (premium license), and The Noun Project[11], so that participants could select their preferred icon.

Using the **user-selected** icons (e.g., Figure 6a), 36 calendar notifications, each set comprising *icon-augmented notifications* and their corresponding (transformed) *text notifications* (see Figure 7) were designed similar to *study 2* (sec 6.2). Half of them represented *density one*, and the other half represented *density two*.

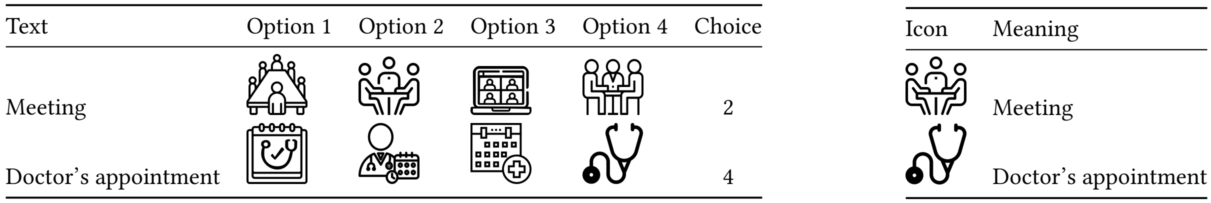
7.3 Study Design and Procedure

A repeated-measures design was used to investigate participants' performances on primary and secondary tasks for the two notification *format*s and two *densities*. Thus, the experiment consisted of four testing blocks, $text_{one}$, $text_{two}$, $icon_{one}$, and $icon_{two}$, each with a duration of four minutes. The blocks were counterbalanced using a Latin square.

7.3.1 Procedure. After briefing the participants on the study and collecting their consent online, they were asked to select their preferred icon out of four choices (e.g., Figure 6a), and if none of the given choices matched their preferences, participants were allowed to suggest new icons. Based on their preference, an icon-to-text mapping (e.g., Figure 6b) was generated for each participant. The entire survey took 10-15 minutes and was done a day before the in-lab study, which allowed the experimenter to prepare materials accordingly and provided participants additional time to get familiarised with their chosen icons. It is important to note that, as shown in Figure 6b, the **icon-to-text mapping** had both icon stimuli used in *icon-augmented notifications* and the corresponding text stimuli used in *text notifications*; thus, thus, the additional time benefits not only the icon conditions but both conditions equally.

During the in-lab experiment, verbal recognition tests were first carried out to remove unfamiliar icons, as was implemented in

[11]https://thenounproject.com/



(a) Icon selection for given texts (here, 2^{nd} and 4^{th} icons are selected) **(b) Corresponding icon-to-text mapping**

Figure 6: User's icon selection and creation of icon-to-text mapping. Icon sources: Flaticon website (premium license) and The Noun Project (by IconTrack, ProSymbols).

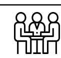
density	regular *text notification*	*icon-augmented notification*		transformed *text notification*
one	Meeting at 4 pm		4 pm	Meeting 4 pm
two	Doctor's appointment in 2 hours		2 hrs	Doctor's appointment 2 hrs
one★	Car is arriving in 5 minutes		5 min	Car 5 min
two★	Mothers' day next week		1 wk	Mothers' day 1 wk

Figure 7: Mapping from *icon-augmented notifications* to transformed *text notifications.* ★ shows the *text reduction* from the 'regular' *text notification* to 'transformed' *text notification.* Icon sources: Flaticon website (premium license) and Google Material Icons.

study 1 (sec 5.4.1). Following that, a training session with no notification and *density one* notification conditions was carried out until participants were confident with the apparatus. Then participants engaged with the four testing blocks, filled out questionnaires, took breaks, and participated in the post-interview, similar to *study 1*. Each participant performed the in-lab experiment in one session lasting 60-80 minutes.

7.3.2 Measures. The same measures from *study 1* (sec 5.5) were used. Additionally, to quantify the perceived differences in the two notification formats due to icons' varying levels of *density*, *Noticeability*: 'How easy or difficult was it to notice the notification?' [80] (under *Reaction*, Table 1) and *Understandability*: 'Once you notice the notification, how easy or difficult was it to understand what it stands for?' [80] (under *Comprehension*) were measured using 7-point Likert scales where 1 = very difficult and 7 = very easy.

7.4 Study 3: Results

Each participant received 24 notifications and 256 targets in total during testing blocks. Participants scored a minimum of 2 (out of 6) for recall accuracy at the end of each notification block and yielded more than 67% for the hit rate. Figure 9, Figure 11, and Figure 10 indicate the mean performance for different measures (see Appendix C.1, **Table 5** for details).

7.4.1 Analysis. Factorial repeated measures ANOVAs or factorial repeated measures ANOVAs after Aligned Rank Transform (ART [102]) when ANOVA assumptions were violated, were applied. Additionally, pairwise comparisons, paired-sample t-test or Wilcoxon signed-rank test were used as post-hoc tests, and Bonferroni correction was applied to p-values for multiple comparisons. Before each test, ANOVA assumptions were tested, and interview data were thematically analyzed, similar to *study 1* (sec 5.6.1).

7.4.2 Icon selection. All participants found that icon selection helpful in recognizing and understanding the meaning of the icons. Only one participant suggested new icons as he could not find four matching icons from the given choices. Based on the survey results and the post-survey interview, all participants chose their icons primarily based on familiarity. Twenty participants also considered the clarity and simplicity of icons, while nine considered the pleasantness of icons. In comparison, two participants preferred detailed and skeuomorphic icons, two participants preferred thin outlines, and three preferred thick outlines. This indicates that people select icons based on icon properties as well as their personal aesthetic tastes.

Surprisingly, a massive disparity between the researcher-selected icons in *study 2* and user-selected icons in this study was found as only 16% of the selected icons matched. For **each icon**, the percentage of participants who selected the researcher-selected icons in *study 2* varied between 4% to 75%. Similarly, for **each participant**, the percentage of user-selected icons that matched the researcher-selected icons in *study 2* varied between 8% and 53%. These results indicate individual differences in icon preferences and the lack of consensus on specific icons due to different mobile platform usage, and possible unfamiliarity issues that existed in *study 2*. For example, Figure 8 shows an example where <email icon> selection was agreed upon by most participants, while three different icons were selected as <meeting icon>.

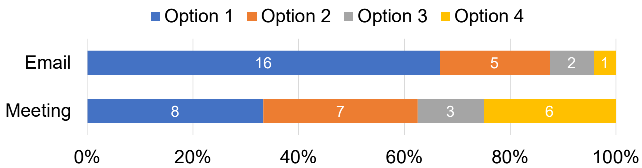
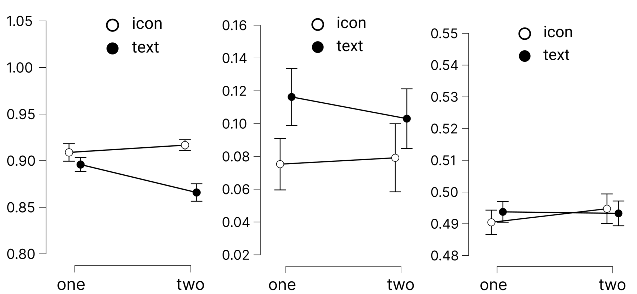


Figure 8: Icon selection preferences for <email icon> and <meeting icon> in *study 3* (N=24). The majority preferred *Option 1* for <email icon>, while user selection for <meeting icon> varied between *Option 1, 2, & 4*. Here *Option 4* represents the researcher-selected icon in *study 2*.

7.4.3 Primary (vigilance) task performance. *Pictogram format* had a significantly higher ($p < 0.05$) hit rate, while the false alarm rate and reaction time were comparable to the *text format* ($p > 0.05$). Moreover, the *density two* degraded the primary task sustainment of the *text format*.

- *Hit rate* (Figure 9a): Repeated-measures ANOVA after ART revealed a significant main effect of *format* ($F_{1,69} = 11.689$, $p = 0.001$, $\eta_p^2 = 0.404$, large effect [50]) and interaction effect ($F_{1,69} = 3.955$, $p = 0.049$, $\eta_p^2 = 0.187$, large effect), but no significant main effect of *density*. Moreover, there were significant ($p < 0.001$) simple main effects for *text format* and *density two*. Post-hoc analyses revealed that the *pictogram format* ($M = 0.913$, $SD = 0.066$) had significantly higher ($p_{bonf} = 0.001$, $r = 0.698$, large effect) mean hit rate than the *text format* ($M = 0.881$, $SD = 0.081$); and $icon_{one}$ and $icon_{two}$ had significantly higher ($p_{bonf} < 0.05$) hit rate than $text_{two}$. This also indicates that when an icon can represent multiple words, *pictogram format* can outperform *text format*.
- *False alarm rate* (Figure 9b): There were no significant effects.
- *Reaction time* (Figure 9c): There were no significant effects.

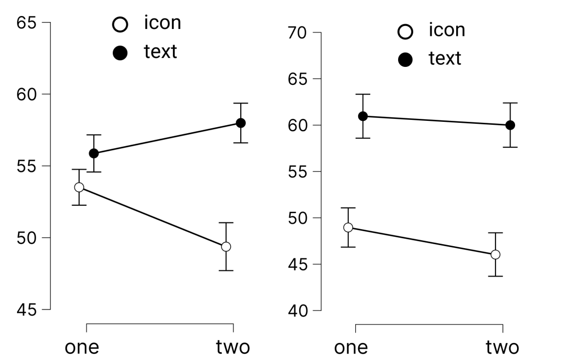


(a) Hit rate **(b) False alarm rate** **(c) Reaction time**

Figure 9: Primary task performance. The X-axis represents the *density*, and the error bars represent the standard error. See Appendix C.1, Table 5 for details.

7.4.4 Interruption. *Icon-augmented notifications* led to a lower cognitive load and perceived interruption than transformed *text notifications*.

- *Unweighted NASA-TLX* (Figure 10a): Repeated-measures ANOVAs showed a significant main effect of *format* ($F_{1,23} = 8.164$, $p = 0.009$, $\eta_p^2 = 0.262$, large effect), but no significant main effect of *density* or interaction effect. Post-hoc analysis showed that the *pictogram format* ($M = 51.44$, $SD = 15.22$) had significantly lower task load ($p_{bonf} < 0.01$, $d = 0.374$, small effect) than the *text format* ($M = 56.93$, $SD = 13.99$). The results for individual indices are given in Appendix C.2.
- *Perceived interruption* (Figure 10b): There was a significant main effect of *format* ($F_{1,23} = 19.068$, $p < 0.001$, $\eta_p^2 = 0.453$, large effect) and interaction effect ($F_{1,23} = 7.261$, $p = 0.013$, $\eta_p^2 = 0.240$, large effect), but no significant main effect of *density*. Post-hoc analyses showed that the *pictogram format* ($M = 47.50$, $SD = 18.28$) had significantly lower interruption ($p_{bonf} < 0.001$, $d = 0.699$, medium effect) than the *text format* ($M = 60.48$, $SD = 18.54$). Furthermore, $icon_{two}$ had significantly lower perceived interruption ($p_{bonf} < 0.05$) than $text_{one}$ and $text_{two}$.
- *Distraction ranking*: All participants, except two (92%) mentioned that the *text format* was most distracting since text required a longer time to read and understand and covered more space on the OHMD. Moreover, twelve participants who recognized the difference between the two *text formats*, mentioned that $text_{two}$ caused the highest distraction. The remaining two participants felt that icons were more distracting as they needed to recall icon meanings to understand *pictogram format* compared to *text format*.



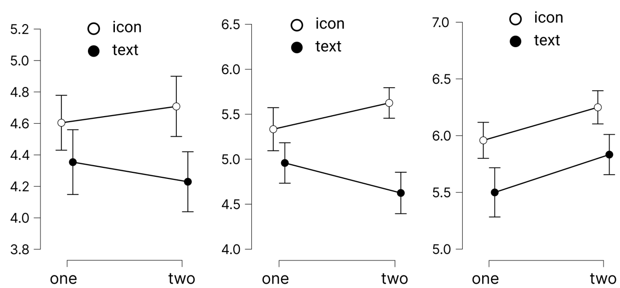
(a) Unweighted NASA-TLX **(b) Perceived interruption**

Figure 10: *Interruption*. The X-axis represents the *density*, and the error bars represent the standard error. See Appendix C.1, Table 5 for details.

7.4.5 Comprehension. Overall, there was no significant difference between *formats* for immediate recall accuracy ($p = 0.055$), but the *pictogram format* had a significantly higher understandability ($p = 0.020$) than the *text format*. There were no main effects of *density* or interaction effect for any measures.

- *Recall Accuracy* (Figure 11a): There was no significant difference ($F_{1,69} = 3.796$, $p = 0.055$) between *pictogram format* ($M = 4.66$, $SD = 1.22$) and *text format* ($M = 4.29$, $SD = 1.15$).
- *Understandability* (Figure 11b): There was only a significant main effect of *format* ($F_{1,69} = 5.667$, $p = 0.020$, $\eta_p^2 = 0.404$, large effect), and the post-hoc analysis indicated *pictogram*

format (M = 5.48, SD = 1.15) had significantly higher (p_{bonf} < 0.05, r = 0.486, large effect) mean value than the *text format* (M = 4.79, SD = 1.75). This indicates that the *pictogram format* was easier to understand during the dual-task scenario than the *text format* when users engaged in an attention-demanding primary task.
- *Understandability ranking*: Fourteen participants (58%) mentioned that the *pictogram format* was easier to understand, easier to interpret, took less time to read, and was more concise. Out of the fourteen, six participants mentioned that the *text format* required more *"mental effort"* to understand than the *pictogram format* while switching their attention between primary and secondary tasks. Four participants mentioned that they felt no difference between the two formats, while the remaining six participants said that the *text format* was easier to understand since it was unambiguous and did not require any recalling of icons.



(a) Recall accuracy (b) Understandability (c) Noticeability

Figure 11: *Comprehension* and *Reaction*. The X-axis represents the *density*, and the error bars represent the standard error. See Appendix C.1, Table 5 for details.

7.4.6 Reaction. In contrast to *study 2*, there was a difference in noticeability levels between the two formats.

- *Noticeability* (Figure 11c): There was only a significant main effect of *format* ($F_{1,69}$ = 8.084, p = 0.006, η_p^2 = 0.252, large effect), and the post-hoc analysis indicated the the *pictogram format* (M = 6.10, SD = 1.08) had a significantly higher (p_{bonf} < 0.05, r = 0.553, large effect) mean value than the *text format* (M = 5.67, SD = 1.19).
- *Noticeability ranking*: Fourteen participants (58%) felt that both the *text* and *pictogram format*s were equally noticeable; seven felt that the *pictogram format* was more noticeable since icons were more eye-catching, while the remaining three participants felt that the *text format* was more noticeable since they were longer and covered more screen space.

7.4.7 Satisfaction. All participants except two (92%) preferred the *pictogram format* over the *text format* as they felt that icons were shorter, easier to recognize and understand, and occluded their OHMD less. Moreover, six of them also noted that, unlike text, icons were recognizable due to their shape, even while they focused on the primary task. The remaining two participants mentioned that, even though both formats caused considerable distractions, the *text format* was easier to interpret accurately.

7.5 Discussion

The results show that *pictogram format* sustained the primary task performance in terms of hit rate, particularly when *pictogram format* had *density two* compared to *text format*. Moreover, independent of *density*, *pictogram format* reduced the *Interruption* while maintaining higher *Comprehension* and *Reaction*.

Considering the results of this study and *study 2*, it can be concluded that user-selected icons improved familiarity, which in turn enhanced multitasking performance. Although, compared to *Study 2*, participants received additional time in this study to become familiar with both icons and texts through repeated self-exposure [86], user selection is the most likely factor contributing to these results. First, participants' selection was based on familiarity, as detailed in sec 7.4.2. Second, as Shen et al. [86] have found, it takes considerable time to become familiar with icons, even with repeated exposure (e.g., more than 3 days with more than 30 minutes of exposure each day). Thus, the survey duration in this study may not have been long enough to achieve a high level of familiarity through repeated self-exposure alone. Furthermore, it can be concluded that the advantages of transforming *text format* to *pictogram format* do indeed depend on replacing texts with icons, as well as the icon familiarity. Hence, users should be given a choice to select icons for *icon-augmented notifications* to minimize interpretative ambiguities; however, we note that it would be challenging to cater to individual preferences. Similarly, when comparing this study with *study 1*, it can be concluded that making the text more concise also reduces the interruption of OHMD notifications.

8 STUDY 4: COMPARE *ICON-AUGMENTED NOTIFICATIONS* AND TRANSFORMED *TEXT NOTIFICATIONS* IN REALISTIC SETTINGS FOR *USER-SELECTED* ICONS

To determine whether the results obtained in the laboratory settings can be generalized to the real world, a realistic qualitative study involving both mobile and stationary settings was conducted. This study was designed similarly to *study 3* (sec 7), except for its tasks and procedure. Thus, the same calendar notifications used in *study 3* (sec 7.2) were used in this study.

8.1 Participants

Twelve volunteers (6 females, 6 males, mean age = 26.6 years, SD = 3.5) participated in the study. They had similar backgrounds to the participants of *study 1* (sec 5.1).

8.2 Apparatus

Our system (a tablet computer) pushed one notification per minute onto the OHMD to avoid flooding participants with too many notifications and provide sufficient notifications to experience the differences between the two notification formats. This also simulated a situation in which users were engaged in a remote conversation [6]. Notifications were randomly displayed with a minimum gap of 40 seconds between each. The random presentation mimics situations where users forget their upcoming events or the events' notifications are not created by the user, such as shared calendars. The notifications were chosen from $text_{one}$, $text_{two}$, $icon_{one}$, $icon_{two}$

with equal probabilities such that they alternated between *text format* and *pictogram format*.

8.3 Tasks

8.3.1 Mobile setting: Navigation. Following the approach of Lucero and Vetek [59], a 1 km outdoor route on the university premises, which consisted of a shared pedestrian and bicycle trail (with road crossings) and a 0.2 km indoor route, was chosen (see Appendix D.1). Participants were familiar with this route and walked the route (primary task) while wearing the OHMD and attended to (pre-selected) calendar notifications (secondary task).

8.3.2 Stationary setting: Browsing. Internet browsing of the participants' choice [9] was chosen as the primary task in this setting and was conducted over 10 minutes on a desktop computer in a lab, while participants also attended to calendar notifications (secondary task) on OHMD.

8.4 Procedure

The briefing of the participant, icon selection, and training were conducted similar to *study 3* (sec 7.3.1); however, participants engaged in the training session while walking in a university lab. As the natural light and weather conditions vary, the time of the experiment was randomly assigned to increase the generalizability of the results. Thus, participants completed the study during the daytime (9 am-6 pm), where they first performed the navigation task and then the browsing task.

Participants walked the predefined route at their comfortable pace while wearing the OHMD. The OHMD was equipped with removable lens shades during the navigation task to ensure the visibility of notifications in outdoor environments. The experimenter walked a few meters behind, carrying the system that automatically pushes notifications to the OHMD. Moreover, to ensure the participants' safety, the experimenter closely monitored participants and stopped them from engaging in dangerous behaviors such as jaywalking. Once participants completed the navigation task (16-20 minutes), they ranked the formats based on distraction, noticeability, understandability, and preference.

Subsequently, participants engaged in the browsing task (8-12 minutes) and indicated their preferences for the stationary setting. In the end, the experimenter conducted a semi-structured interview to capture the reasons for each choice and understand how participants attended to notifications during the mobile and stationary settings. Each participant completed the study in one session lasting 50-65 minutes.

8.5 Study 4: Results

Some differences in the perception of notifications for mobile and stationary settings were found. The codes resulting from the thematic analysis were grouped based on the task differences and measures used in *study 3*.

8.5.1 Interruption. During the navigation task, six participants felt that the *text format* was more distracting, especially for longer text, while the remaining participants did not feel any difference between the two notification formats.

During the browsing task, six participants felt that both formats had similar distraction levels, five participants felt that *text format* was more distracting due to primary task interference, while the remaining participant felt that *pictogram format* was more distracting due to primary task interference.

8.5.2 Comprehension. Overall, there were no significant differences between the *format*s for immediate recall accuracy, but *text format* showed a higher understandability ranking during the browsing task.

- *Recall Accuracy*: During both tasks, participants remembered[12] *icon-augmented notifications* (navigation task: M = 45.2%, SD = 19.1%; browsing task: M = 69.3%, SD = 8.1%) more than *text notifications* (navigation task: M = 34.5%, SD = 17.2%; browsing task: M = 64.0%, SD = 15.6%) even though there was no statistically significant difference.
- *Understandability ranking*: During the navigation task, five participants felt that both formats were equally understandable, four participants felt that the *text format* was more understandable, while the remaining three participants felt that the *pictogram format* was more understandable.
  Similarly, during the browsing task, six participants felt that both formats were equally understandable, five participants felt that the *text format* was more understandable, while the remaining participant felt that the *pictogram format* was more understandable.
  Five participants who recognized the varying *density* of text format also noted that *"short texts and icons [notifications] were [the] same [easy to understand]. Still, longer texts were harder to understand"*.

8.5.3 Reaction. Results for the navigation task contrasted with those of the previous studies. Six participants felt that the *text format* was more noticeable due to its longer length and more light emission from OHMD, four participants felt that both formats were equally noticeable, while the remaining two participants felt that the *pictogram format* was more eye-catching.

With regards to the browsing task, eight participants felt that both formats were equally noticeable; three participants shared that the *pictogram format* was more noticeable since icons were distinct from the browsing text, while the remaining participant felt *text format* was more noticeable as it occluded the browsing content more.

As expected, outdoor lighting conditions affected the visibility of digital content on OHMD [26, 29, 34] which affected notification noticeability. All participants felt that notifications were more noticeable indoors than outdoors, and four of them felt notifications were less visible with a bright or green background due to the lack of contrast differences.

8.5.4 Satisfaction. Overall, most participants preferred the *pictogram format* regardless of their task.

During the navigation task, nine participants preferred the *pictogram format*, quoting similar reasons to *study 3*; shorter, easy to read, less occlusion, and less distraction than the *text format*.

[12]Here, *Recall Accuracy* = $\frac{\text{No of correctly recalled notifications}}{\text{No of total seen notifications}} \times 100$, which account for the differences in participants' walking speeds

Two participants shared equal preferences for both formats, and the remaining participant preferred the *text format*.

During the browsing task, all except one participant preferred the *pictogram format* while the remaining participant preferred the *text format*; however, as expected, participants mentioned that their preference depended on their familiarity with icons.

8.6 Discussion

The real-life settings yielded similar results to the lab-controlled ones (*study 1, and 3*) for *Interruption* (perceived cost of interruption) and *Comprehension* (for *Recall Accuracy*). However, the *Reaction* (e.g., noticeability) was affected by lighting conditions, and *Comprehension* (for perceived understandability) was affected by the complexity of the primary task. Specifically, outdoors, when external lighting became high, *text format* and its longer length became more noticeable than *pictogram format*.

Moreover, primary tasks during realistic settings were less attention-demanding; thus, more attentional resources were available to attend to the notifications than in lab settings which supports the finding that *text format* was perceived as more understandable than *pictogram format* during the browsing setting. A comparison of this result with that of *study 3* suggests that when the primary task's complexity increased or needed higher attentional control, users familiar with icons could understand *pictogram format* better than *text format*.

Lastly, five participants felt that their OHMDs shook as they physically moved, which affected the legibility of the two notification formats. All of the participants felt that long texts' legibility was more prone to shaking, while one participant felt that the *pictogram format*'s legibility was also prone to shaking.

9 GENERAL DISCUSSION

Through *study 1*, we verified that transforming *text notifications* to *icon-augmented notifications* can reduce the interruption of calendar OHMD notifications and improve primary task performance. In *studies 2 and 3*, we found that the advantages of *icon-augmented notifications* depended on the users' familiarity with icons, the number of words that the icons can represent, and the extent of text reduction. In lab settings, *icon-augmented notifications* transformed from *text notifications* were equally comprehensible and noticeable; they were also less interruptive and preferred. *Study 4* confirmed that these results could be largely generalized to real-world situations, albeit with limitations in terms of noticeability. Overall, it can be concluded that incorporating icons can indeed minimize interruption during multitasking without compromising the reaction and comprehension of OHMD notifications. However, to achieve such benefits, the *icon-augmented notifications* need to be carefully designed according to some influencing factors; such as: icon familiarity, encoding density, and external brightness.

9.1 Reasons behind the disparity

The findings of our study suggest a plausible reason behind the observed disparity in literature (sec 2.3). Pictograms are likely to show advantages if they are highly familiar to the users and have an encoding density greater than 1; otherwise, they are likely to show no advantage or generate worse performance than text. This phenomenon is reflected in previous studies. Tanveer et al. [90] reported that participants found text feedback to be more effective and easier to learn than bar-like pictorial feedback on OHMDs during public speaking. In their study design, only one-word text was used; thus, the encoding density was 1. In addition, since the bar charts used in the experiments are not frequently used in everyday life, users may not be familiar. Thus, the text condition is likely to yield better performance. Similarly, Warnock et al. [98, 99] found no difference between the text and icon formats. In their work, the information contained in the text and icon-based notifications was limited to a single word, which equals an encoding density of 1. This finding is in line with the conclusion in *study 3* that when the encoding density of icons is 1, both *formats* have a similar level of distraction for *user-selected* icons.

On the other hand, most studies favoring pictograms typically have a relatively small set of well-defined, commonly used pictograms with relatively high encoding density (much higher than 1, often close to 2 or beyond). For example, Ells and Dewar [27] compared traffic signs with their corresponding pictograms and found that the results significantly favor pictograms. Upon a closer examination of the stimuli used, we found that their study one compared eight pictograms with eight text messages. These pictograms were well defined and easy to understand with an average encoding density of 1.88, while their study two had fourteen pictograms with an average encoding density of 2.07. Similar analysis can be applied to Kline et al.'s [52] study as well, where they used four pictograms with an average encoding density of 1.75.

9.2 When and how to use *icon-augmented notifications*

With a better understanding of the conditions in which *icon-augmented notifications* can have advantages, we can now discuss when and how to use them.

First, *icon-augmented notifications* are less suitable for delivering general-purpose notifications. This is because the vocabulary of general-purpose notifications is not restricted, making it difficult to find suitable icons for the diverse meaning a notification wants to express. Even if it is possible to find many icons, remembering them will be challenging for users, and it is unlikely that they will perform better than pure text-based notifications.

Nevertheless, *icon-augmented notifications* can be used in specialized domains; such as: notifying users of their personalized calendar or reminder events. Most of the calendar/reminder events are routine and recurring [51, 95]; thus, they can be easily represented using a few well-designed icons. Moreover, given the personalized context of calendars, as many of the events are set by the users themselves, it is also much easier for them to understand the meaning when they see a reminder notification.

In **Figure 12**, we suggest an approach to convert calendar events to *icon-augmented notifications* on OHMDs. Users first identify the frequent, recurring, or important tasks/events from their personal calendar, and for each event, the system (or user) identifies the primary information. The system then provides a set of icons for users to choose from or allows users to suggest/modify icons if necessary. Then, they can either directly go to *step 4* to create *icon-augmented notifications* or use an intermediate scaffolding step (*step*

3) where both text and icons are used in a redundant fashion to help users to learn the associations first, then transit to the more concise version of *icon-augmented notifications* later.

In addition to calendar events, OHMDs are used in medical, navigation, and manufacturing domains [67, 97], where users need higher levels of attentional control. In such specialized domains, it is easier to define a set of well-designed icons (e.g., [10]) that become highly familiar to users through repeated use. In such cases, we envision that *icon-augmented notifications* can ease the notification handling on OHMDs. Nevertheless, there are a number of additional considerations when applying *icon-augmented notifications* in practical applications.

9.2.1 Multiple secondary info. In this study, the content of the notifications was categorized into *primary info* and *secondary info* to transform the *text format* to *pictogram format*; yet, this transformation may not be suitable for complex text notifications with multiple secondary information. Research on text illustrations has explored the use of an abstraction called 'meaning space' to map text with illustrations [21, 30]. Such literature primarily focuses on identifying suitable images to illustrate text fragments or vice versa. By incorporating the guidelines and algorithms used in the text illustration literature, different transformation techniques can be evaluated and implemented with existing applications to identify the optimal transformations for different categories of notifications.

9.2.2 Shared interpretation. Furthermore, when pictograms are used in communication applications, both senders and receivers should have a shared understanding of the meaning of pictograms and sufficient context for interpretation [18]. This may also apply to *icon-augmented notifications* when the creator/sender and receiver are different, such as messenger notifications. However, each party may build a shared understanding of *icon-augmented notifications* with usage over time to overcome this limitation, similar to emoticons [105].

9.2.3 Using icon-augmented notifications for mobile OHMD usage. From *study 4*, we identify that *icon-augmented notifications* are easier to perceive in suboptimal conditions during mobile OHMD usage [27, 100, Ch 6]. Specifically, even though virtual content on OHMD is blurred when users focus on the physical world, i.e., limited OHMD focal distance [47, 48], the pictogram format on OHMD can be easier to recognize than the text format. This is because shapes have a larger visual detection angle than text [47, 104]. Moreover, longer text notifications require users to use their central vision to read word by word [100, Ch 6], while *icon-augmented notifications* can be read using the central vision in a single glance, reducing recognition effort and time.

9.2.4 One issue related to mobile OMHD usage is external brightness. To overcome the external brightness, designs can increase the salience of *icon-augmented notifications* in outdoor environments. For instance, borders can outline notifications (e.g., Figure 13), or colors can contrast/blend with the environment (e.g., [4, 25]).

9.2.5 Generalizing results to other OHMDs. Compared to advanced OHMDs, like Microsoft HoloLens2 (HL2)[13] or Nreal Light (Nreal)[14], which have a larger FoV (field of view), use 3D content, and support various anchoring techniques (HL2 and Nreal supports the head, body, and world anchoring), the OHMD prototype we used, BT-300, has a smaller FoV, uses 2D content, and supports head anchoring. Given that the features of BT-300 are a subset of those from HL2 and Nreal, those HL2 and Nreal that are using similar configurations to the BT-300 can more easily replicate our results obtained on BT-300. If the more advanced OHMDs use features that are specific to their capabilities, e.g., world anchoring, given the limited world anchoring distance (e.g., recommended distance for HL2 is 1.25m - 5m [15]) [48], we believe our results still largely hold, as previous studies comparing pictograms and text-based traffic signs from the similar physical distance (e.g., [27]) showed favorable results towards pictograms due to the high encoding density of icons (sec 9.1).

10 LIMITATIONS

In this study, calendar-related notifications with **one** *secondary info* were investigated and showed that *icon-augmented notifications* are an effective alternative for *text notifications* in the OHMD context. However, considering the limited number of icons compared to text, extending the use of *icon-augmented notifications* to other types of notifications (e.g., messenger notifications), especially with **multiple** *secondary info* should be done with care as results may not apply evenly to all types of notifications.

As discussed in *study 4* (sec 8.6), external brightness mainly affected the noticeability of *icon-augmented notifications*, and shaking mainly affected the legibility of *text notifications*. With the advancement in technology, such as retinal projection [48] (e.g., Vaunt glasses[16]), the effects of external brightness will be minimized, and the use of fonts/icons which are less susceptible to shakiness [61] can minimize legibility issues. Although the selected vigilance task in *Studies 1-3* can mimic dynamic conditions in realistic situations (sec 5.3), it did not simulate severe conditions, such as bumping into someone while walking in a crowded street. It also did not include scenarios involving potential danger that one can encounter in real-life augmented reality usage (e.g., reduced depth of focus and reaction time [81]). Similarly, even though *Study 4* verified the ecological validity during walking, it did not test the many possible real-world scenarios in which people can engage in other tasks while attending notifications (e.g., conversing [80], reading and walking obstacle-rich environments [96]). Thus, further validation in a variety of realistic scenarios can further enhance the ecological validity of our results. However, we believe the *pictogram format* will have a higher salience during shakes, be easier to perceive in sub-optimal conditions (sec 9.2.3), and provide higher attention control (sec 5.7, sec 9.2); thus, offering more advantages during those scenarios over *text format*.

Furthermore, we only considered the notification content to isolate the effects of text and pictorial formats; however, in the event that we add additional elements such as app sources, which are currently present on mobile phone notifications (e.g., Figure 1a) [3, 5], the visualization of the notification needs to be re-arranged.

[13]https://www.microsoft.com/en-us/hololens/hardware
[14]https://www.nreal.ai/light/
[15]https://learn.microsoft.com/en-us/windows/mixed-reality/design/comfort
[16]https://www.theverge.com/2018/2/5/16966530/intel-vaunt-smart-glasses-announced-ar-video

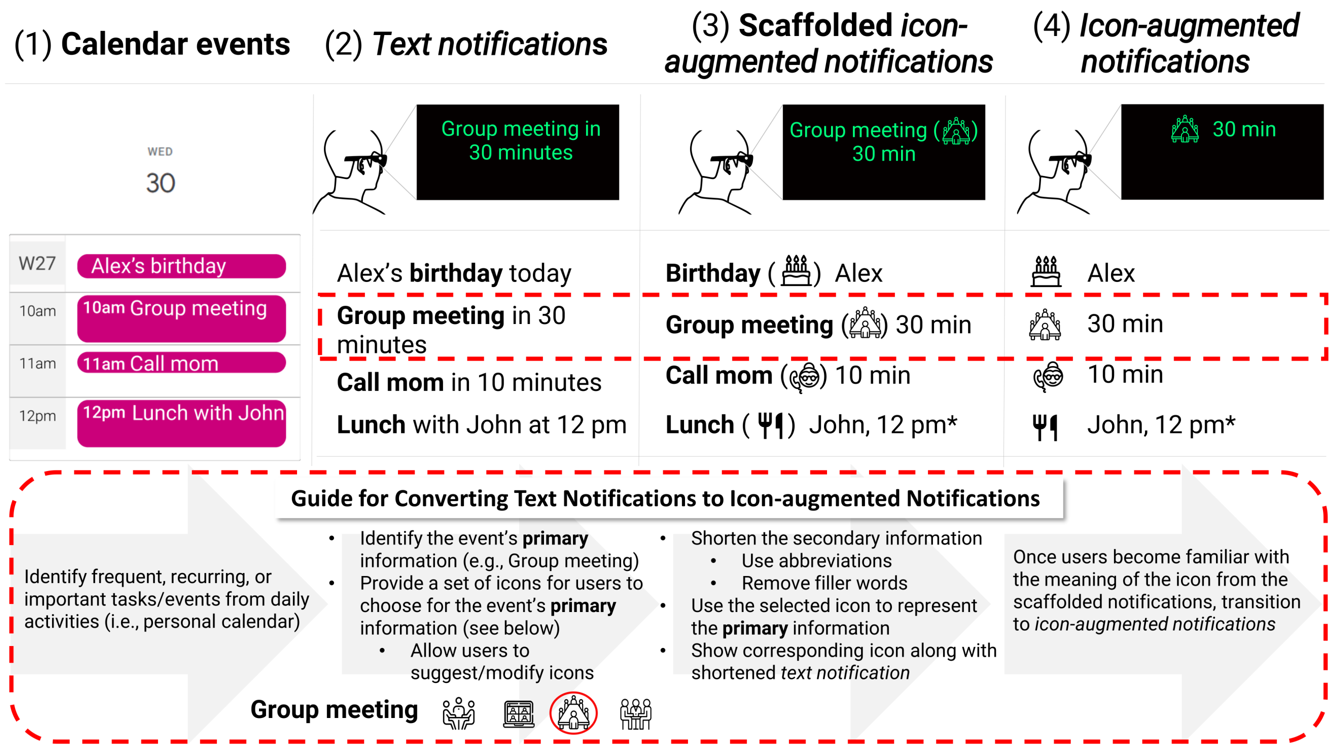


Figure 12: The above illustration describes the guidelines for converting OHMD *text notifications* to *icon-augmented notifications* and applying the guidelines to convert calendar events into *icon-augmented notifications*. The group meeting event is highlighted as an elaborated example. Icon sources: Flaticon website (premium license), Google Material Icons, and The Noun Project (by Matt Brooks). Note: * indicates a scenario where there is more than one *secondary info* that is not investigated in this paper.

Current OHMDs notifications (e.g., Google Glasses [36]) use the same layout elements as notifications on mobile phones to preserve consistency, though this may not be optimal for all OHMDs, which is outside this study's scope.

Finally, we also note that our studies only examined notifications for short durations for tech-savvy participants in limited realistic scenarios. Although our sample sizes are moderate [12], given the medium and large effect sizes, the results are generalizable to similar populations. However, the results do not capture long-term effects and may not evenly apply to other populations, such as older adults and people with visual impairments.

11 CONCLUSION

We proposed that transforming text notifications to *icon-augmented notifications* can reduce the interruption of OHMD notifications and enable better multitasking when designed properly. Using three controlled experiments, we demonstrated that *icon-augmented notifications* are advantageous on OHMDs, as pictograms can richly convey meanings of multiple texts, and the text content can be shortened. In addition, two plausible reasons (i.e., icon familiarity and encoding density) were identified for the observed disparity in literature (sec 2.3) related to the effective use of pictograms in notifications. Using a realistic setting, results from controlled experiments were applicable to realistic situations but with notable constraints when using an outdoor setting.

Considering the inherent properties of pictograms (sec 2.3,9.2.3), we believe that our results on *icon-augmented notifications* are generalizable to visual notifications in other computing devices, although we only verified them for OHMDs. Furthermore, we believe *icon-augmentation* can be applied to other digital information presentations outside notifications which still need further investigation.

Future work can be geared towards transforming more complex *text notifications* to *pictogram formats*, considering the effects of diverse backgrounds and capabilities of various OHMDs with the help of eye-tracking, which will enable the widespread and mainstream use of pictograms in notifications.

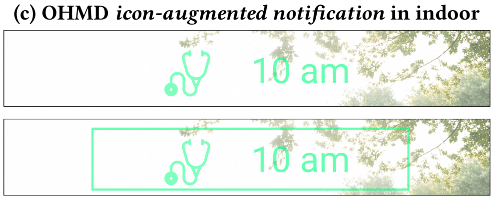
ACKNOWLEDGMENTS

This research is supported by the National Research Foundation, Singapore, under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-016). It is also supported in part by the Ministry of Education, Singapore, under its MOE Academic Research Fund Tier 2 programme (MOE-T2EP20221-0010), and by a research grant #22-5913-A0001 from the Ministry of Education of Singapore. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views

(a) OHMD *text notification* in indoor

(b) OHMD *text notification* in outdoor

(c) OHMD *icon-augmented notification* in indoor



(d) OHMD *icon-augmented notification* in outdoor without border (top) or with border (bottom)

Figure 13: Effects of external brightness/lighting on the noticeability of OHMD notifications. In indoor situations without bright external light, both *text* and *pictogram formats* have similar noticeability and legibility (*a* vs. *c*). But in outdoor conditions with bright external light, the longer *text format* is more noticeable than, the more concise *pictogram format* (*b* vs. *d − top*). Thus, it's recommended to increase the visual footprint of the *icon-augmented notification* (e.g., add a border to increase its noticeability, *d − bottom*). Note: This figure may need to be viewed in color to see the differences more clearly. Icon source: Flaticon website (premium license).

of the National Research Foundation or the Ministry of Education, Singapore.

We would like to express our gratitude to Felicia Tan for her generous help with proofreading the paper, and to the volunteers who participated in our studies. We also thank the reviewers for their valuable time and insightful comments, which helped to improve this paper.

REFERENCES

[1] Piotr D. Adamczyk and Brian P. Bailey. 2004. If not now, when?: the effects of interruption at different moments within task execution. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*. ACM Press, Vienna, Austria, 271–278. https://doi.org/10.1145/985692.985727

[2] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (July 2018), 1–27. https://doi.org/10.1145/3214261

[3] Android. 2021. Android Notifications Overview. https://developer.android.com/guide/topics/ui/notifiers/notifications Retrieved April 8, 2021 from https://developer.android.com/guide/topics/ui/notifiers/notifications.

[4] Siao Ying Ang and Giap Weng Ng. 2020. A Study on the Effect of the Real-world Backgrounds using Colour Blending Technique on Optical See-through AR User Interface Design. In *2020 6th International Conference on Interactive Digital Media (ICIDM)*. 1–4. https://doi.org/10.1109/ICIDM51048.2020.9339664

[5] Apple. 2022. Notifications - System experiences - Components - Human Interface Guidelines - Design - Apple Developer. https://developer.apple.com/design/human-interface-guidelines/components/system-experiences/notifications/

[6] Daniel Avrahami, Susan R. Fussell, and Scott E. Hudson. 2008. IM waiting: timing and responsiveness in semi-synchronous communication. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*. ACM Press, San Diego, CA, USA, 285. https://doi.org/10.1145/1460563.1460610

[7] Brian P Bailey, Joseph A Konstan, and John V Carlis. 2001. The Effects of Interruptions on Task Performance, Annoyance, and Anxiety in the User Interface. *International Conference on Human-Computer Interaction* (2001), 593–601.

[8] Magdalena Bartłomiejczyk. 2013. Text and image in traffic signs. *LINGUISTICA SILESIANA* (2013), 111–131. https://journals.pan.pl/publication/103135/edition/89148/ Publisher: Polska Akademia Nauk • Oddział w Katowicach.

[9] Thomas Beauvisage. 2009. Computer Usage in Daily Life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 575–584. https://doi.org/10.1145/1518701.1518791

[10] Nigel Bevan and Jonathan Earthy. 2018. Benefiting from ISO standards. *The Wiley Handbook of Human Computer Interaction* 1 (2018), 51–69.

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[12] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[13] Monica J. Camacho, Bruce A. Steiner, and Barry L. Berson. 1990. Icons vs. Alphanumerics in Pilot-Vehicle Interfaces. *Proceedings of the Human Factors Society Annual Meeting* 34, 1 (1990), 11–15. https://doi.org/10.1177/154193129003400104

[14] Steve Caplin. 2001. *Icon design: Graphic icons in computer interface design*. Watson-Guptill Publications, Inc.

[15] Cándida Castro, Tim Horberry, and Francisco Tornay. 2004. The Effectiveness of Transport Signs. In *The Human Factors of Transport Signs*. CRC Press.

[16] Isha Chaturvedi, Farshid Hassani Bijarbooneh, Tristan Braud, and Pan Hui. 2019. Peripheral vision: a new killer app for smart glasses. In *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19*. ACM Press, Marina del Ray, California, 625–636. https://doi.org/10.1145/3301275.3302263

[17] C. M. Chewar, D. Scott McCrickard, and Alistair G. Sutcliffe. 2004. Unpacking critical parameters for interface design: evaluating notification systems with the IRC framework. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (DIS '04)*. Association for Computing Machinery, New York, NY, USA, 279–288. https://doi.org/10.1145/1013115.1013155

[18] Heeryon Cho, Toru Ishida, Toshiyuki Takasaki, and Satoshi Oyama. 2008. Assisting Pictogram Selection with Semantic Interpretation. In *The Semantic Web: Research and Applications (Lecture Notes in Computer Science)*, Sean Bechhofer, Manfred Hauswirth, Jörg Hoffmann, and Manolis Koubarakis (Eds.). Springer, Berlin, Heidelberg, 65–79. https://doi.org/10.1007/978-3-540-68234-9_8

[19] Soon Hau Chua, Simon T. Perrault, Denys J. C. Matthies, and Shengdong Zhao. 2016. Positioning Glass: Investigating Display Positions of Monocular Optical See-Through Head-Mounted Display. In *Proceedings of the Fourth International Symposium on Chinese CHI - ChineseCHI2016*. ACM Press, San Jose, USA, 1–6. https://doi.org/10.1145/2948708.2948713

[20] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. 2016. Workspace Awareness in Collaborative AR using HMDs: A User Study Comparing Audio and Visual Notifications. In *Proceedings of the 7th Augmented Human International Conference 2016 on - AH '16*. ACM Press, Geneva, Switzerland, 1–8. https://doi.org/10.1145/2875194.2875204

[21] Filipe Coelho and Cristina Ribeiro. 2012. Image Abstraction in Crossmedia Retrieval for Text Illustration. In *Advances in Information Retrieval*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Ricardo Baeza-Yates, Arjen P. de Vries, Hugo Zaragoza, B. Barla Cambazoglu, Vanessa Murdock, Ronny Lempel, and Fabrizio Silvestri (Eds.). Vol. 7224. Springer Berlin Heidelberg, Berlin, Heidelberg, 329–339. https://doi.org/10.1007/978-3-642-28997-2_28

[22] Enrico Costanza, Samuel A. Inverso, Elan Pavlov, Rebecca Allen, and Pattie Maes. 2006. Eye-q: Eyeglass peripheral display for subtle intimate notifications. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services - MobileHCI '06*. ACM Press, Helsinki, Finland, 211. https://doi.org/10.1145/1152215.1152261

[23] Ed Cutrell, Mary Czerwinski, and Eric Horvitz. 2001. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance.

In *INTERACT 2001*. IOS Press, 263–269. https://www.microsoft.com/en-us/research/publication/notification-disruption-and-memory-effects-of-messaging-interruptions-on-memory-and-performance/

[24] Ionut Damian, Chiew Seng (Sean) Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting Social Interactions: Realtime Behavioural Feedback using Social Signal Processing Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 565–574. https://doi.org/10.1145/2702123.2702314

[25] Juan David Hincapie-Ramos, Levko Ivanchuk, Srikanth Kirshnamachari Sridharan, and Pourang Irani. 2014. SmartColor: Real-time color correction and contrast for optical see-through head-mounted displays. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Munich, Germany, 187–194. https://doi.org/10.1109/ISMAR.2014.6948426

[26] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. 2014. Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. *IEEE Transactions on Visualization and Computer Graphics* 20, 1 (Jan. 2014), 125–139. https://doi.org/10.1109/TVCG.2013.86

[27] Jerry G. Ells and Robert E. Dewar. 1979. Rapid Comprehension of Verbal and Symbolic Traffic Sign Messages. *Human Factors* 21, 2 (April 1979), 161–168. https://doi.org/10.1177/001872087902100203

[28] Epson. 2021. Epson Moverio BT-300 Smart Glasses. https://www.epson.com.au/products/ProjectorAccessories/Moverio_BT-300_Specs.asp Retrieved September 27, 2021 from https://www.epson.com.au/products/ProjectorAccessories/Moverio_BT-300_Specs.asp.

[29] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Gregory F. Welch. 2020. Exploring the Limitations of Environment Lighting on Optical See-Through Head-Mounted Displays. In *Symposium on Spatial User Interaction (SUI '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3385959.3418445

[30] Ali Farhadi, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier, and David Forsyth. 2010. Every Picture Tells a Story: Generating Sentences from Images. In *Computer Vision – ECCV 2010*, Kostas Daniilidis, Petros Maragos, and Nikos Paragios (Eds.). Vol. 6314. Springer Berlin Heidelberg, Berlin, Heidelberg, 15–29. https://doi.org/10.1007/978-3-642-15561-1_2

[31] Niaja Farve, Tal Achituv, and Pattie Maes. 2016. User Attention with Head-Worn Displays. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM Press, Santa Clara, California, USA, 2467–2473. https://doi.org/10.1145/2851581.2892530

[32] S. Fitrianie and L.J.M. Rothkrantz. 2005. Communication in Crisis Situations Using Icon Language. In *2005 IEEE International Conference on Multimedia and Expo*. 1370–1373. https://doi.org/10.1109/ICME.2005.1521685

[33] Shogo Fukushima, Takeo Hamada, and Ari Hautasaari. 2020. Comparing World and Screen Coordinate Systems in Optical See-Through Head-Mounted Displays for Text Readability while Walking. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Porto de Galinhas, Brazil, 649–658. https://doi.org/10.1109/ISMAR50242.2020.00093

[34] Joseph L. Gabbard, J. Edward Swan, and Deborah Hix. 2006. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence* 15, 1 (Feb. 2006), 16–32. https://doi.org/10.1162/pres.2006.15.1.16

[35] Jean Dickinson Gibbons and Subhabrata Chakraborti. 2020. *Nonparametric Statistical Inference*. CRC Press.

[36] Google. 2021. User Interface | Glass Explorer Edition. https://developers.google.com/glass/design/ui Retrieved May 8, 2021 from https://developers.google.com/glass/design/ui.

[37] Mark A Goss-Sampson. 2019. Statistical Analyzis in JASP: A Guide for Students. (2019), 123.

[38] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[39] Eric Horvitz, Johnson Apacible, and Muru Subramani. 2005. Balancing Awareness and Interruption: Investigation of Notification Deferral Policies. In *User Modeling 2005 (Lecture Notes in Computer Science)*, Liliana Ardissono, Paul Brna, and Antonija Mitrovic (Eds.). Springer, Berlin, Heidelberg, 433–437. https://doi.org/10.1007/11527886_59

[40] Eric Horvitz, Paul Koch, and Johnson Apacible. 2004. BusyBody: creating and fielding personalized models of the cost of interruption. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 507–510. https://doi.org/10.1145/1031607.1031690

[41] Peter S. Houts, Cecilia C. Doak, Leonard G. Doak, and Matthew J. Loscalzo. 2006. The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling* 61, 2 (May 2006), 173–190. https://doi.org/10.1016/j.pec.2005.05.004

[42] Sheng-Cheng Huang, Randolph G. Bias, and David Schnyer. 2015. How are icons processed by the brain? Neuroimaging measures of four types of visual stimuli used in information systems. *Journal of the Association for Information Science and Technology* 66, 4 (2015), 702–720. https://doi.org/10.1002/asi.23210

[43] Carl J Huberty and John D Morris. 1992. Multivariate analysis versus multiple univariate analyses. (1992).

[44] Katri Hämeen-Anttila, Kati Kemppainen, Hannes Enlund, J Bush Patricia, and Airaksinen Marja. 2004. Do pictograms improve children's understanding of medicine leaflet information? *Patient Education and Counseling* 55, 3 (Dec. 2004), 371–378. https://doi.org/10.1016/j.pec.2003.04.006

[45] Shamsi T. Iqbal and Eric Horvitz. 2010. Notifications and awareness: a field study of alert usage and preferences. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10*. ACM Press, Savannah, Georgia, USA, 27. https://doi.org/10.1145/1718918.1718926

[46] Sarah J. Isherwood, Siné J. P. McDougall, and Martin B. Curry. 2007. Icon Identification in Context: The Changing Role of Icon Characteristics With User Experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49, 3 (June 2007), 465–476. https://doi.org/10.1518/001872007X200102

[47] Yoshio Ishiguro and Jun Rekimoto. 2011. Peripheral vision annotation: noninterference information presentation method for mobile augmented reality. In *Proceedings of the 2nd Augmented Human International Conference on - AH '11*. ACM Press, Tokyo, Japan, 1–5. https://doi.org/10.1145/1959826.1959834

[48] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-mounted Displays. *Comput. Surveys* 54, 6 (July 2021), 120:1–120:36. https://doi.org/10.1145/3453157

[49] Nuwan Janaka, Chloe Haigh, Hyeongcheol Kim, Shan Zhang, and Shengdong Zhao. 2022. Paracentral and near-peripheral visualizations: Towards attention-maintaining secondary information presentation on OHMDs during in-person social interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3502127

[50] Matthew Kay. 2021. Effect Sizes with ART. https://cran.r-project.org/web/packages/ARTool/vignettes/art-effect-size.html Retrieved July 19, 2022 from https://cran.r-project.org/web/packages/ARTool/vignettes/art-effect-size.html.

[51] J. F. Kelley and Alphonse Chapanis. 1982. How professional persons keep their calendars: Implications for computerization. *Journal of Occupational Psychology* 55, 4 (1982). https://doi.org/10.1111/j.2044-8325.1982.tb00098.x

[52] Theresa J. Babbitt Kline, Laura M. Ghali, Donald W. Kline, and Steven Brown. 1990. Visibility Distance of Highway Signs among Young, Middle-Aged, and Older Observers: Icons Are Better than Text. *Human Factors* 32, 5 (Oct. 1990), 609–619. https://doi.org/10.1177/001872089003200508

[53] Elisa Maria Klose, Nils Adrian Mack, Jens Hegenberg, and Ludger Schmidt. 2019. Text Presentation for Augmented Reality Applications in Dual-Task Situations. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 636–644. https://doi.org/10.1109/VR.2019.8797992

[54] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. 2016. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1011–1020. https://doi.org/10.1145/2858036.2858359

[55] Robert S. Laramee and Colin Ware. 2002. Rivalry and interference with a head-mounted display. *ACM Transactions on Computer-Human Interaction* 9, 3 (Sept. 2002), 238–251. https://doi.org/10.1145/568513.568516

[56] Felix Lauber and Andreas Butz. 2014. In-your-face, yet unseen?: improving head-stabilized warnings to reduce reaction time. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 3201–3204. https://doi.org/10.1145/2556288.2557063

[57] Luis Leiva, Matthias Böhmer, Sven Gehring, and Antonio Krüger. 2012. Back to the app: the costs of mobile application interruptions. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 291–294. https://doi.org/10.1145/2371574.2371617

[58] Cesar Leos-Toro, Geoffrey T. Fong, Samantha B. Meyer, and David Hammond. 2019. Perceptions of effectiveness and believability of pictorial and text-only health warning labels for cannabis products among Canadian youth. *International Journal of Drug Policy* 73 (Nov. 2019), 24–31. https://doi.org/10.1016/j.drugpo.2019.07.001

[59] Andr\'{e}s Lucero and Akos Vetek. 2014. NotifEye: using interactive glasses to deal with notifications while walking in public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology - ACE '14*. ACM Press, Funchal, Portugal, 1–10. https://doi.org/10.1145/2663806.2663824

[60] Kris Luyten, Donald Degraen, Gustavo Rovelo Ruiz, Sven Coppers, and Davy Vanacken. 2016. Hidden in Plain Sight: an Exploration of a Visual Language for Near-Eye Out-of-Focus Displays in the Peripheral View. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 487–497. https://doi.org/10.1145/2858036.

2858339

[61] Yuki Matsuura, Tsutomu Terada, Tomohiro Aoki, Susumu Sonoda, Naoya Isoyama, and Masahiko Tsukamoto. 2019. Readability and legibility of fonts considering shakiness of head mounted displays. In *Proceedings of the 23rd International Symposium on Wearable Computers*. ACM, London United Kingdom, 150–159. https://doi.org/10.1145/3341163.3347748

[62] Gerard McAtamney and Caroline Parker. 2006. An examination of the effects of a wearable display on informal face-to-face communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 45–54. https://doi.org/10.1145/1124772.1124780

[63] D.Scott McCrickard, Richard Catrambone, C.M Chewar, and John T Stasko. 2003. Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. *International Journal of Human-Computer Studies* 58, 5 (May 2003), 547–582. https://doi.org/10.1016/S1071-5819(03)00022-3

[64] D. Scott McCrickard and C. M. Chewar. 2003. Attuning notification design to user goals and attention costs. *Commun. ACM* 46, 3 (March 2003), 67. https://doi.org/10.1145/636772.636800

[65] D. Scott McCrickard, C. M. Chewar, Jacob P. Somervell, and Ali Ndiwalana. 2003. A model for notification systems evaluation—assessing user goals for multitasking activity. *ACM Transactions on Computer-Human Interaction (TOCHI)* 10, 4 (Dec. 2003), 312–338. https://doi.org/10.1145/966930.966933

[66] Siné J. P. Mcdougall, Martin B. Curry, and Oscar de Bruijn. 1999. Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers* 31, 3 (Sept. 1999), 487–519. https://doi.org/10.3758/BF03200730

[67] Stefan Mitrasinovic, Elvis Camacho, Nirali Trivedi, Julia Logan, Colson Campbell, Robert Zilinyi, Bryan Lieber, Eliza Bruce, Blake Taylor, David Martineau, Emmanuel L. P. Dumont, Geoff Appelboom, and E. Sander Connolly Jr. 2015. Clinical and surgical applications of smart glasses. *Technology and Health Care* 23, 4 (Jan. 2015), 381–401. https://doi.org/10.3233/THC-150910

[68] Terhi Mustonen, Mikko Berg, Jyrki Kaistinen, Takashi Kawai, and Jukka Häkkinen. 2013. Visual Task Performance Using a Monocular See-Through Head-Mounted Display (HMD) While Walking. *Journal of experimental psychology: applied* 19, 4 (2013), 12. https://doi.org/10.1037/a0034635

[69] Federica Nenna, Marco Zorzi, and Luciano Gamberini. 2021. Augmented Reality as a research tool: investigating cognitive-motor dual-task during outdoor navigation. *International Journal of Human-Computer Studies* 152 (Aug. 2021), 102644. https://doi.org/10.1016/j.ijhcs.2021.102644

[70] Eyal Ofek, Shamsi T. Iqbal, and Karin Strauss. 2013. Reducing disruption from subtle information delivery during a conversation: mode and bandwidth investigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3111–3120. https://doi.org/10.1145/2470654.2466425

[71] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*. ACM Press, Portland, Oregon, USA, 919. https://doi.org/10.1145/1054972.1055101

[72] H. Pashler. 1994. Dual-task interference in simple tasks: data and theory. *Psychological Bulletin* 116, 2 (Sept. 1994), 220–244. https://doi.org/10.1037/0033-2909.116.2.220

[73] Celeste Lyn Paul, Anita Komlodi, and Wayne Lutters. 2015. Interruptive notifications in support of task management. *International Journal of Human-Computer Studies* 79 (July 2015), 20–34. https://doi.org/10.1016/j.ijhcs.2015.02.001

[74] Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51, 1 (Feb. 2019), 195–203. https://doi.org/10.3758/s13428-018-01193-y

[75] Martin Pielot, Karen Church, and Rodrigo de Oliveira. 2014. An in-situ study of mobile phone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*. ACM Press, Toronto, ON, Canada, 233–242. https://doi.org/10.1145/2628363.2628364

[76] Ashwin Ram and Shengdong Zhao. 2021. LSVP: Towards Effective On-the-go Video Learning Using Optical Head-Mounted Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 30:1–30:27. https://doi.org/10.1145/3448118

[77] Pei-Luen Patrick Rau and Jian Zheng. 2019. Modality capacity and appropriateness in multimodal display of complex non-semantic information stream. *International Journal of Human-Computer Studies* 130 (Oct. 2019), 166–178. https://doi.org/10.1016/j.ijhcs.2019.06.008

[78] Thomas S Redick, Zach Shipstead, Matthew E Meier, Janelle J Montroy, Kenny L Hicks, Nash Unsworth, Michael J Kane, D Zachary Hambrick, and Randall W Engle. 2016. Cognitive Predictors of a Common Multitasking Ability: Contributions From Working Memory, Attention Control, and Fluid Intelligence. *Journal of Experimental Psychology. General* 145, 11 (Nov. 2016), 1473–1492. https://doi.org/10.1037/xge0000219

[79] Javier Roca, Beatriz Insa, and Pilar Tejero. 2018. Legibility of Text and Pictograms in Variable Message Signs: Can Single-Word Messages Outperform Pictograms? *Human Factors* 60, 3 (May 2018), 384–396. https://doi.org/10.1177/0018720817751623

[80] Rufat Rzayev, Susanne Korbely, Milena Maul, Alina Schark, Valentin Schwind, and Niels Henze. 2020. Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, Tallinn Estonia, 1–11. https://doi.org/10.1145/3419249.3420095

[81] Eric E. Sabelman and Roger Lam. 2015. The real-life dangers of augmented reality. *IEEE Spectrum* 52, 7 (July 2015), 48–53. https://doi.org/10.1109/MSPEC.2015.7131695

[82] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-scale assessment of mobile notifications. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 3055–3064. https://doi.org/10.1145/2556288.2557189

[83] Valerio Santangelo, Paola Finoia, Antonino Raffone, Marta Olivetti Belardinelli, and Charles Spence. 2008. Perceptual load affects exogenous spatial orienting while working memory load does not. *Experimental Brain Research* 184, 3 (Jan. 2008), 371–382. https://doi.org/10.1007/s00221-007-1108-8

[84] Martin Sarter and John P. Bruno. 2002. Vigilance. In *Encyclopedia of the Human Brain*, V. S. Ramachandran (Ed.). Academic Press, New York, 687–699. https://doi.org/10.1016/B0-12-227210-2/00357-5

[85] L. Sevens. 2018. *Words Divide, Pictographs Unite: Pictograph Communication Technologies for People with an Intellectual Disability*. Ph. D. Dissertation. https://lirias.kuleuven.be/retrieve/518329

[86] Zhangfan Shen, Chengqi Xue, and Haiyan Wang. 2018. Effects of Users' Familiarity With the Objects Depicted in Icons on the Cognitive Performance of Icon Identification. *i-Perception* 9, 3 (April 2018), 2041669518780807. https://doi.org/10.1177/2041669518780807 Publisher: SAGE Publications.

[87] David Shinar and Margreet Vogelzang. 2013. Comprehension of traffic signs with symbolic versus text displays. *Transportation Research Part F: Traffic Psychology and Behaviour* 18 (May 2013), 72–82. https://doi.org/10.1016/j.trf.2012.12.012

[88] Jacob Somervell, C M Chewar, and D Scott McCrickard. 2002. Evaluating graphical vs. textual secondary displays for information notification. *Proceedings of the ACM Southeast Conference, Raleigh NC* (2002), 153–160.

[89] Cary Stothart, Ainsley Mitchum, and Courtney Yehnert. 2015. The attentional cost of receiving a cell phone notification. *Journal of Experimental Psychology: Human Perception and Performance* 41, 4 (Aug. 2015), 893–897. https://doi.org/10.1037/xhp0000100

[90] M. Iftekhar Tanveer, Emy Lin, and Mohammed (Ehsan) Hoque. 2015. Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15*. ACM Press, Atlanta, Georgia, USA, 286–295. https://doi.org/10.1145/2678025.2701386

[91] Channary Tauch and Eiman Kanjo. 2016. The Roles of Emojis in Mobile Phone Notifications. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 1560–1565. https://doi.org/10.1145/2968219.2968549

[92] John Theios and Paul C. Amrhein. 1989. Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review* 96 (1989), 5–24. https://doi.org/10.1037/0033-295X.96.1.5 Place: US Publisher: American Psychological Association.

[93] Charles Tijus, Javier Barcenilla, Brigitte Cambon de Lavalette, and Jean-Guy Meunier. 2007. The Design, Understanding and Usage of Pictograms. *Written Documents in the Workplace* (Jan. 2007), 17–31. https://doi.org/10.1163/9789004253254_003

[94] Maciej Tomczak and Ewa Tomczak. 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in sport sciences* 1, 21 (2014), 19–25.

[95] Manas Tungare, Manuel Perez-Quinones, and Alyssa Sams. 2008. An Exploratory Study of Calendar Use. *arXiv:0809.3447 [cs]* (Sept. 2008). http://arxiv.org/abs/0809.3447

[96] Kristin Vadas, Nirmal Patel, Kent Lyons, Thad Starner, and Julie Jacko. 2006. Reading on-the-go: a comparison of audio and hand-held displays. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services (MobileHCI '06)*. Association for Computing Machinery, New York, NY, USA, 219–226. https://doi.org/10.1145/1152215.1152262

[97] X. Wang, S. K. Ong, and A. Y. C. Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (March 2016), 1–22. https://doi.org/10.1007/s40436-015-0131-4

[98] David Warnock. 2011. A subjective evaluation of multimodal notifications. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. 461–468. https://doi.org/10.4108/

icst.pervasivehealth.2011.246001

[99] David Warnock, Marilyn McGee-Lennon, and Stephen Brewster. 2013. Multiple notification modalities and older users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1091–1094. https://doi.org/10.1145/2470654.2466139

[100] Christopher D. Wickens, Justin G. Hollands, Simon Banbury, and Raja Parasuraman. 2015. *Engineering psychology and human performance* (4th ed.). Pearson Education.

[101] Susan Wiedenbeck. 1999. The use of icons and labels in an end user application program: An empirical study of learning and retention. *Behaviour & Information Technology* 18, 2 (Jan. 1999), 68–82. https://doi.org/10.1080/014492999119129

[102] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[103] Krzysztof Wołk, Agnieszka Wołk, Krzysztof Marasek, and Wojciech Glinkowski. 2017. Pictogram-based mobile first medical aid communicator. *Procedia Computer Science* 121 (2017), 3–10. https://doi.org/10.1016/j.procs.2017.11.002

[104] Michelle Yeh, Christopher D. Wickens, and F. Jacob Seagull. 1999. Target Cuing in Visual Search: The Effects of Conformality and Display Location on the Allocation of Visual Attention. *Human Factors* 41, 4 (Dec. 1999), 524–542. https://doi.org/10.1518/001872099779656752

[105] Rui Zhou, Jasmine Hentschel, and Neha Kumar. 2017. Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 748–759. https://doi.org/10.1145/3025453.3025800

A STUDY 1

A.1 Calendar notifications

Table 3 represents the calendar notifications used in *study 1*.

A.2 *Recall Accuracy* calculation

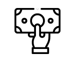
Consider a hypothetical case where a participant sees two notifications '<meeting, 4 pm>' and '<birthday, tomorrow>'. If the participant recalled only one notification, '<meeting, tomorrow>' (e.g., the participant wrote 'meeting is on tomorrow' in the questionnaire), they would get only 0.5 points since 'tomorrow' (*secondary info*) is not the correct secondary info for 'meeting'(*primary info*). Since six notifications were displayed during testing blocks, participants could score a maximum of 6 points for *Recall Accuracy* in each *text notification* or *icon-augmented notification format*.

A.3 Individual NASA-TLX Indices

Overall *icon-augmented notifications* had a lower task load than *text notifications*, as shown in Figure 14. A repeated-measures ANOVA showed significant main effects of *notification format* on the overall score ($F_{2,28} = 45.076$, $p < 0.001$) as well as all individual indices ($p < 0.001$), except for *Physical Demand*. A post-hoc analysis with Bonferroni correction revealed that *no-notification* yielded significantly lower ($p < 0.001$) task load results than *text notification* and *icon-augmented notification* for all measures except *Physical Demand*.

Furthermore, *icon-augmented notification* was significantly lower ($p < 0.10$) than *text notification* for overall score and *Frustration*. On all indices, including overall score (Figure 14), the sorted order of task-load from lower to higher was; *no-notification* < *icon-augmented notification* < *text notification*.

Table 3: Twenty-four calendar notifications used in *study 1* with their *text format* and *pictogram format*. These are adapted from real mobile-phone notifications. Icon sources: Google Material Icons and Flaticon website (premium license). Each icon's source is attached as a hyperlink to the icon itself.

text format	*pictogram format*	
Meeting at 4 pm		4 pm
Doctor's appointment in 2 hours		2 hrs
Lunch with Lee		Lee
Birthday party tomorrow		1 d
Visitor coming on Friday		Friday
Car is arriving in 5 minutes		5 min
Email meeting agenda		agenda
Delivery in 3 days		3 d
Pay $100		$100
Credit card bill today		today
Presentation at noon		12 pm
Pay rental on Monday		Monday
text format	*pictogram format*	
Buy milk and eggs tonight		tonight
Exercise in 40 minutes		40 min
Check flight status		status
Reply Alex		Alex
Coffee break at 3 pm		3 pm
Renew driving license		renew
Backup computer tonight		tonight
Movie on Friday		Friday
Download the e-bill		e-bill
Cycling at 6 pm		6 pm
Call Mary		Mary
Valentine day in 2 weeks		2 wk

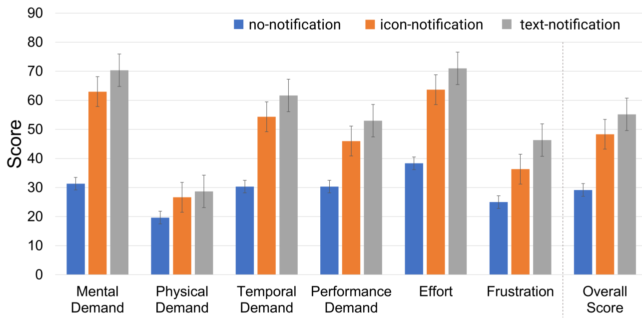


Figure 14: NASA-TLX scores for *no-notification*, *icon-augmented notification*, and *text notification* in *Study 1* (N=15). Error bars represent standard errors.

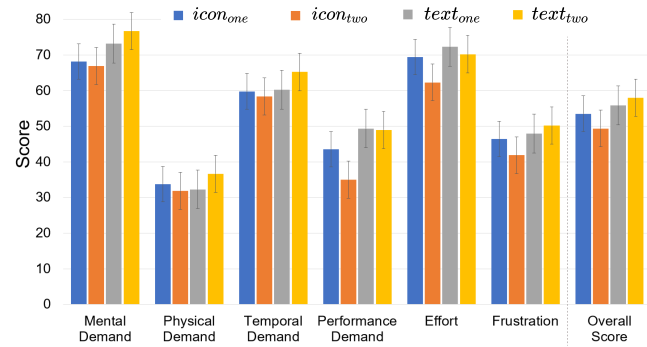


Figure 15: NASA-TLX scores for $icon_{one}$, $icon_{two}$, $text_{one}$, and $text_{two}$ in *Study 3* (N=24). Error bars represent standard errors.

B STUDY 2

B.1 Quantitative data

The mean performance measures of *study 2* are presented in Table 4. Surprisingly, contrary to *study 1*, *pictogram format* yielded a lower *hit rate* and similar *reaction time* than *text format*, although none were significant.

C STUDY 3

C.1 Quantitative data

The mean performance measures of *study 3* are presented in Table 5.

C.2 Individual NASA-TLX Indices

Figure 15 shows the individual indices for each *format* × *density* combination.

Overall *pictogram format* had significantly lower task load than *text format*. A repeated-measures ANOVA showed significant main effects ($p < 0.05$) of *format* on the overall score, *Mental Demand*, *Performance Demand*, and *Effort*. With respect to the main effects of *format*, *pictogram format* yielded lower mean values than *text format*.

However, except for *Performance Demand*, there were no significant main effects of *density*. Similarly, there were significant interaction effects ($p < 0.05$) only for overall score and *Mental Demand*.

Figure 16: The *study 4* route (red arrow) includes the outdoors and indoors. Map source: OpenStreetMap

D STUDY 4

D.1 The route

Figure 16 shows the route participants took in *study 4*. Participants went past three bus stops, five vehicle crossings, and a small park during outdoor navigation. During indoor navigation, participants walked through covered buildings.

E PROGRAMMING CODES

Codes for this study can be found at https://github.com/NUS-HCILab/IconAugmentedNotification. If you encounter any issues accessing them, please contact the authors.

Table 4: *Study 2* performance in dual-task scenario (N = 12). Colored bars show the relative value of each measure for different notification formats. Here, *Icon* = *icon-augmented notification* and *Text* = transformed *text notification*.

Measure	*H*		*F*		*RT*		*Recall Accuracy*		*RTLX*		*Perceived Interruption*	
Format	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Icon	0.938	0.080	0.041	0.061	0.481	0.049	4.17	1.04	44.10	14.72	47.92	20.22
Text	0.949	0.045	0.073	0.095	0.481	0.043	4.13	1.31	45.66	13.55	50.54	21.00

Table 5: *Study 3* performance in dual-task scenario (N = 24). Colored bars show the relative value of each measure for different *format* × *density* combinations.

Measure	*H*		*F*		*RT*	
Format	M	SD	M	SD	M	SD
$icon_{one}$	0.909	0.064	0.075	0.109	0.490	0.037
$icon_{two}$	0.917	0.069	0.079	0.108	0.495	0.055
$text_{one}$	0.896	0.068	0.116	0.130	0.494	0.039
$text_{two}$	0.866	0.092	0.103	0.134	0.493	0.040

Measure	*Recall Accuracy*		*Noticeability*		*Understandability*		*RTLX*		*Perceived Interruption*	
Format	M	SD	M	SD	M	SD	M	SD	M	SD
$icon_{one}$	4.60	1.25	5.96	1.12	5.33	1.20	53.51	14.60	48.96	17.88
$icon_{two}$	4.71	1.21	6.25	1.03	5.63	1.10	49.38	15.85	46.04	18.94
$text_{one}$	4.35	1.26	5.50	1.38	4.96	1.71	55.87	14.83	60.96	18.40
$text_{two}$	4.23	1.06	5.83	0.96	4.63	1.81	57.99	13.33	60.00	19.05